First Course in

METRIC SPACES

B. K. Tyagi

FOUNDATION®
BOOKS

Delhi • Bangalore • Mumbai • Kolkata • Chennai • Hyderabad • Pune

Preface

Real analysis is the branch of mathematical analysis which is primarily concerned with limit processes and continuity of real valued functions defined on the subsets of the real line. The considerations of these notions have led mathematicians to generalise these in various ways, for example, to study the convergence of sequences of real or complex numbers, the continuity of complex valued functions defined over the domains of complex numbers, etc. The reader who has done a course of real analysis must have noted that these concepts of limit, continuity of a function and convergence of sequence depend for their meaning on the notion of the absolute value of the difference of two numbers when they are regarded as the points on the real line, which is nothing but the distance between the points. It is no exaggeration if one says that the notion of distance permeates everywhere in the entire course of real analysis and also in geometry.

When the above notion of distance (or metric) for real numbers is extended to abstract sets in general, then the resulting mathematical structures are known as metric spaces. The metric spaces were first introduced by Maurice Fréchet in his thesis for doctorate in mathematics in 1906 and are being investigated since then. The metric spaces provide an appropriate setting to study 'closeness' of objects by defining a distance between them. This notion of distance captures the semantics of the real life usage of the meaning of the word 'distance'. The metric spaces have evolved further, principally, as topological spaces in which continuity is freed from the notion of distance. The topological spaces will, however, be outside the scope of this book. The class of metric spaces is a subclass of the class of all topological spaces.

Though the book is self-contained, for the appreciation of the various topics, the knowledge of a first level course in real analysis is an added advantage. The book contains standard material drawn from various texts cited in the bibliography and has been written primarily for undergraduate students. It is hoped that the book will be useful to teachers and research scholars in this area as well.

The need for such a book arises in view of the fact that a course in metric spaces is offered at graduate level in universities throughout the world. The book is completely rigorous, and an effort has been made to give the motivation of the various notions. The book provides ample opportunities to the reader to fill in his/her reasoning in the gaps left in proofs of theorems and examples. The book contains many exercises, and the reader is encouraged to solve them.

The book is organised as follows: Chapter 1 contains the preliminary material required for other chapters of the book. It contains the discussion on sets, relations, functions (mappings) predicates and quantifiers, order relations, number system, the axiom of choice, cardinal numbers, ordinal numbers, linear algebra, and some basic concepts of real analysis that describes the convergence of sequences. An effort is made to put in place all the basic notions and notations in a concise and explicit manner, that are used later on. The section 1.13 on ordinal numbers is not used in the chapters that follow, although it is important on its own. It is included here to get a kind of completion in our so called 'Counting procedures' which continue in Section 1.12 and extend to Section 1.13. A beginner may not go through all the preliminaries at one go. He/She may skip the Sections 1.11-1.13 on the first reading and return to these when required while studying the main text from Chapter 2 onwards.

Chapter 2 begins with the formal definition of a metric space. Several examples are given to show the richness of the scope of the definition. This chapter contains the basic concepts that are defined in metric spaces. The class of metric spaces is a subclass of the class of all pseudometric spaces discussed in Section 2.14. This chapter also contains the procedures to obtain the new metric spaces from the given ones in the form of subspaces (Section 2.13) and product spaces (Section 2.15). The famous Cantor set is described in Section 2.17.

Chapter 3 generalises the notion of convergence of a sequence to metric spaces. The cluster point of a sequence is defined. Here the notion of convergence of a sequence is also discussed in the product of metric spaces.

Chapter 4 deals with limit and continuity of mappings defined from one metric space to another metric space. Since continuous mappings are reasonably well-behaved theoretically, and are important from the point of view of applications, the reader will come across a very detailed examination of the notion of continuity. The notion of uniform continuity of mappings which is stronger than the notion of continuity is also examined. The chapter also contains the discussion of uniform convergence of sequences of mappings in certain metric spaces. The lower and upper semi-continuous mappings are introduced in Section 4.9.

Chapter 5 gives a very detailed analysis of the notion of completeness. It begins with the proofs of completeness of several metric spaces. Several equivalent characterisations of completeness are obtained. The completion of a metric space is discussed in Section 5.4. Baire's Category Theorem which describes an important property of complete metric spaces is a subject matter of Section 5.5. The chapter also includes the behaviour of various types of mappings in relation to complete metric spaces. Section 5.11 obtains Banach spaces of continuous and bounded mappings. Section 5.12 contains a characterisation of completeness in terms of Cauchy filter base. The chapter closes with a property of a G_δ-set in complete metric spaces.

Chapters 6 and 7 cover the notions of compactness and connectedness, respectively, which are at the base of entire modern mathematical analysis. In addition to the discussion of compactness of subsets in various metric spaces. Chapter 6 contains several equivalent characterisations of compactness such as sequential compactness, Bolzano Weierstrass property, countable compactness and completeness and total boundedness. Section 6.9 covers the compact sets in the metric spaces of continuous mappings which are characterised by Ascoli-Arzela theorem. It also contains the applications of this theorem: Peano theorem for the existence of a solution of a differential equation is described here. The local compactness is described in Section 6.10, and the final section describes Peano space filling curve. Chapter 7 gives the description of various types of connectedness and their examples: connected metric spaces, locally connected metric spaces, path connected metric spaces and totally disconnected metric spaces. Further, it describes the maximal connected subsets, called components. Section 7.4 covers the behaviour of continuous mappings in relation to connected metric spaces. The chapter closes with a description of non-homeomorphic metric spaces.

The famous Weierstrass approximation theorem and Stone-Weierstrass approximation theorem are the subject matter covered in Chapter 8. These theorems approximate the continuous real or complex valued mappings. The chapter closes with an application that shows the existence of a continuous everywhere and differentiable nowhere mapping.

Chapter 9 includes Banach's fixed point theorem and its applications to prove the existence of a unique solution of: a differential equation (Picard's theorem), Fredholm and Volterra integral equations and a system of linear equations. The implicit function theorem is also described here.

Finally, Chapter 10 describes the Cauchy's construction of real numbers as the equivalence classes of Cauchy sequences of rational numbers. It is shown that real numbers form a complete ordered field, and that any two complete ordered fields are isomorphic.

I am indebted to Dr J.K. Kohli, a dedicated topologist, teaching at Hindu College, University of Delhi who readily spared his time to discuss many concepts and examples on innumerable occasions. I also express my gratitude to Professor Sharfuddin Ahmad, Jamia Millia Islamia, New Delhi and Dr Naseem Ajmal, Zakir Hussain College, University of Delhi for their encouragement. I wish to acknowledge my sense of appreciation due to Central Library, University of Delhi for providing the source material and the working space at the research floor. However, I am alone responsible for errors and oversights in this text, and I will gratefully receive the readers comments and suggestions for the improvement of this book. Last but not the least, I am thankful to Rajeev Arora who has gone through the handwritten manuscript, and typed it with a lot of patience.

B. K. Tyagi

Contents

Preface **v**

1. Preliminaries **1**
1.1 Sets 1
1.2 Algebra of Sets 2
1.3 Arbitrary Unions and Intersections 3
1.4 Relations 4
1.5 Mappings 7
1.6 Predicates, Quantifiers, Negations, Converse and Contrapositive 17
1.7 Partially Ordered Sets and Lattices 21
1.8 Number System 26
1.9 Finite Sets and Infinite Sets 46
1.10 Countable and Uncountable Sets 49
1.11 Axiom of Choice, Well-ordered Sets and Infinite Sets 54
1.12 Cardinal Numbers 60
1.13 Ordinal Numbers 65
1.14 Vector Spaces and Normed Spaces 77
1.15 Real Sequences 82

2. Basic Concepts in Metric Spaces **86**
2.1 Definition and Examples of Metric Spaces 86
2.2 Open Spheres 99
2.3 Neighbourhood of a Point 103
2.4 Bounded Sets 105
2.5 Open Sets 110
2.6 Equivalent Metrics 114
2.7 Closed Sets 118
2.8 Limit Point and Closure of a Set 121
2.9 Adherent Points and Isolated Points 125
2.10 Interior, Boundary and Exterior of a Set 127
2.11 Dense Sets, Perfect Sets and Separability 129
2.12 Second Axiom of Countability 134
2.13 Subspaces 136
2.14 Pseudometric Spaces 139
2.15 Product of Metric Spaces 142
2.16 The Cantor Set 145
2.17 Totally Bounded Sets 149

3. Sequences in Metric Spaces **153**
3.1 The Limit of a Sequence 153
3.2 Cluster Point of a Sequence 156
3.3 Convergence in Products 159
3.4 Characterisation of Equivalent Metrics 162

Chapter 1

Preliminaries

1.1 Sets

Though the reader at undergraduate level have some exposure to set theory, one needs to be alert to the basic definitions in the area. For instance, consider the question: What is the definition of a set? There are alternative formal approaches to introduce the notion of a set. The interested reader is referred to excellent books on set theory by Halmos [23], Franklin [19], Mendelson [32], Enderton [17], among others in Bibliography. We attempt a definition of a set in Appendix. The collections which are used by working mathematicians in their every day life are accepted as sets by all definitions of a set. For example, the collections of: numbers, the voters of a state, the undergraduate students in Delhi University in 1980 are all examples of sets. Some collections of sets are also accepted as sets, but the collection of all sets is not a set.

In this text, the capital letters A, B, ..., X, Y, Z, possibly with suffixes, are used to denote sets of objects. If an object a is a member of a set A, we express this fact by the notation

$$a \in A \tag{1.1}$$

and read it as: "a belongs to A", "a is in A", or "a is member of A". An object in A is also called an element of A or a point in A.

The equality predicate '=' is used throughout the text to mean the logical identity. Thus, $a = b$ if a and b are symbols for the same object. Similarly, $a \neq b$ is used to denote the fact that a and b are symbols for different objects.

A set A is said to be a subset of a set B, written as $A \subseteq B$, if $x \in A$ implies that $x \in B$. Two sets A and B are said to be equal, denoted $A = B$, if $A \subseteq B$ and $B \subseteq A$. For example, if $A =$ set of integers, and $B =$ the set of real numbers, then $A \subseteq B$, B is not a subset of A, and A and B are, of course, not equal, that is $A \neq B$. We say that a set A is a proper subset of B, denoted $A \subset B$ or $A \subsetneqq B$, if $A \subseteq B$ and $A \neq B$.

Most often a set can be identified by a property that is common to all its elements. For example,

$A = \{\ x : x$ is capital of a country $\}$,

$B = \{\ k : k$ is a positive integer divisible by 4 $\}$.

Here the braces $\{\ldots\}$ stands for "set of" and the colon : stands for "such that". Thus, the second of the above examples will read as: B is the set of positive integers that are divisible by 4.

Remark. In the nineteenth century, a set was considered to be a collection of elements having some property P, and clearly is the basis of the notation: $\{\ x : x$ has the property $P\}$. Intuitively, it is difficult to see what possibly could go wrong with this. In 1901, however, Russel made the crucial observation that the above understanding of a set leads to a contradiction: Let A be the set of all sets x such that x is not a member of x, that is

$$A = \{x : x \notin x\}.$$

From this we get the contradiction: $A \in A$ if and only if $A \notin A$. This contradiction is known as the Russel's paradox. Nevertheless, the notation $\{\ x : x$ has the property $P\ \}$ for a set is used extensively in whole of mathematics where we put some control on the properties that can be allowed to form a set.

Consider the set

$$A = \{x : x \text{ is an integer such that } x \text{ is less than 3 and greater than 4}\}.$$

Since, there is no integer less than 3 and greater than 4, the set A does not contain any element. Such a set is called an empty set, and is usually denoted by $\emptyset$. Thus, an empty set $\emptyset$ is a set which contains no element. An empty set is called a null set. An empty set is unique, and therefore we speak of "the empty set". The empty set is a proper subset of every non-empty set.

If a set contains only one element x, it is denoted by $\{x\}$ and called a singleton set.

1.2 Algebra of Sets

The sets are combined in several ways to obtain new sets. The operation of union "$\cup$" applied to two sets A_1 and A_2 puts together the elements of A_1 and A_2. It is like the students of two classes sitting in two different rooms are made to sit in a single room. Formally, let A_1 and A_2 be two sets. The union of A_1 and A_2, denoted by $A_1 \cup A_2$, is defined as

$$A_1 \cup A_2 = \{x : x \in A_1 \quad \text{or} \quad x \in A_2\}.$$

Similarly, if there are n sets $A_1, A_2, \ldots, A_n$, their union, denoted by $\bigcup_{i=1}^{n} A_i$, is defined as

$$\bigcup_{i=1}^{n} A_i = \{x : x \in A_i \text{ for some } i \in \{1, 2, \ldots, n\}\}.$$

The notation $\bigcup_{i=1}^{n} A_i$ is an abbreviation for $A_1 \cup A_2 \cup \ldots \cup A_n$. The union of a finite number of sets is termed as "the finite union".

The operation of intersection when applied to two sets produces the set of common elements in those two sets. Formally, if A_1 and A_2 are two sets, then the intersection of A_1 and A_2, denoted by $A_1 \cap A_2$, is defined as

$$A_1 \cap A_2 = \{x : x \in A_1 \text{ and } x \in A_2\}.$$

The symbol "$\cap$" stands for the intersection. Similarly, if $A_1, A_2, \ldots, A_n$ are sets, their intersection, denoted by $\bigcap_{i=1}^{n} A_i$, is defined as

$$\bigcap_{i=1}^{n} A_i = \{x : x \in A_i \text{ for all } i = 1, 2, \ldots, n\}.$$

Two sets A and B are said to be disjoint, if $A \cap B = \emptyset$.

The operation of Cartesian product on sets is also a way of making new sets from the given ones. The notion of Cartesian product of sets, however, uses another basic concept that is of ordered pairs. The ordered pair of two elements a and b, denoted by (a, b) is defined as the set

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

Thus, an ordered pair (a, b) is the set having as its elements the set $\{a\}$ and the set $\{a, b\}$. The first co-ordinate of (a, b) is a which is common to both $\{a\}$, $\{a, b\}$. The second co-ordinate of (a, b) is b which is in one of the members of the set $\{\{a\}, \{a, b\}\}$. When $a = b$, the ordered pair $(a, a) = \{\{a\}\}$, the set containing the set $\{a\}$. If we assume that $(a_1, b_1) = (a_2, b_2)$, it is true that $a_1 = a_2$ and $b_1 = b_2$, and this is usually the property in which the mathematicians are interested when using ordered pairs.

Now we define the Cartesian product of sets. Given two sets A_1 and A_2, the Cartesian product of A_1 and A_2, denoted by $A_1 \times A_2$, is defined as

$$A_1 \times A_2 = \{(a, b) : a \in A_1 \text{ and } b \in A_2\}.$$

If $A_1 = A_2 = A$, then $A_1 \times A_2$ is denoted by A^2.

The difference of a set A from a set B, denoted by $A - B$, is defined as

$$A - B = \{x : x \in A \text{ and } x \notin B\}.$$

The set $A - B$ is sometimes called the complement of B in A. If X is a set and A is a subset of X, then the complement of A in X is usually denoted by A', that is $A' = X - A$.

If A is a set, the power set of A, denoted by $\mathbb{P}(A)$, is the set of all subsets of A.

If $A = \emptyset$, then the power set of A is the set $\mathbb{P}(A) = \mathbb{P}(\emptyset) = \{\emptyset\}$. The power set of $\mathbb{P}(\emptyset)$ is the set

$$\mathbb{P}(\mathbb{P}(\emptyset)) = \{\emptyset, \{\emptyset\}\}.$$

Similarly, $\mathbb{P}(\mathbb{P}(\mathbb{P}(\emptyset))) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$ and so on.

The following facts, known as De Morgen's laws, are used throughout the text without reference:

1. $A - (B \cup C) = (A - B) \cap (A - C)$,
2. $A - (B \cap C) = (A - B) \cup (A - C)$.

The equality (1) above states that the complement of the union is equal to the intersection of complements, and similarly (2) can be formulated.

1.3 Arbitrary Unions and Intersections

In the preceding section, we have defined the union and intersection of finite number of sets. These operations are extended to arbitrary number of sets.

Let $\mathcal{A}$ be an arbitrary family of sets, that is, set of sets. The union of all sets in $\mathcal{A}$, denoted by $\bigcup_{A\in\mathcal{A}} A$, or by $\bigcup\{A : A \in \mathcal{A}\}$, is defined as

$$\bigcup_{A\in\mathcal{A}} A = \{x : x \in A \text{ for some } A \text{ in } \mathcal{A}\}.$$

As an illustration, if $\mathcal{A}$ is an empty family of sets, that is, a family that contains no set, then $\bigcup_{A\in\mathcal{A}} A$ is the empty set.

The intersection of an arbitrary family of sets $\mathcal{A}$, denoted $\bigcap_{A\in\mathcal{A}} A$ or $\bigcap\{A : A \in \mathcal{A}\}$, is defined as

$$\bigcap_{A\in\mathcal{A}} A = \{x : x \in A \text{ for all } A \text{ in}\mathcal{A}\}.$$

If some A in $\mathcal{A}$ happens to be the empty set, then it is clear that $\bigcap_{A\in\mathcal{A}} A$ is empty. If $\mathcal{A}$ is an empty family, its intersection we continue to denote by $\bigcap_{A\in\mathcal{A}} A$, then the question arises which elements belong to this intersection. If all the sets in $\mathcal{A}$ are subsets of a set X, then we can think of an element in X and whether it is in $\bigcap_{A\in\mathcal{A}} A$; so let x be an element in X. Then x will not be in the intersection $\bigcap_{A\in\mathcal{A}} A$ if there exists a set A in $\mathcal{A}$ such that x is not in A; since $\mathcal{A}$ is empty, there is no such set A in $\mathcal{A}$; therefore, x is in the intersection $\bigcap_{A\in\mathcal{A}} A$. Since x is an arbitrary element in X, it follows that $\bigcap_{A\in\mathcal{A}} A = X$. Thus, the intersection of an empty family of subsets of X is X.

It may be noted that while considering the intersection of an empty family of sets in the preceding paragraph, we were forced to consider a master set X, and all other sets in $\mathcal{A}$ were subsets of X; such a set is called a universal set. In most situations in mathematics, a universal set is involved, and all other sets consist of elements of this universal set.

We remark at this point that phrases "family of sets", "set of sets" and "collection of sets", all have identical meaning.

1.4 Relations

There may be a universe (a set) consisting of objects that are related to the objects in the same universe, or to the objects in several different universes in some way. Mathematics and in fact all fields of knowledge endeavour to make significant statements about these relationships. These relationships at times may be quite intricate. So a quest of a mathematician begins by formulating simple types of relationships in which two universes same or distinct are involved; such a relationship is usually called a binary relation. For example, in the universe of all living human beings A, the relationship "father of" is a binary relation that relates two persons (objects) x and y of A if x is father of y. More specifically, we have the following definition.

Definition 1.4.1 A binary relation R from A to B is a subset of $A \times B$. If $A = B$, we say that R is a (binary) relation on A.

We sometimes write xRy to mean that $(x, y) \in R$.

Example 1.4.2

1. Let $A = B = \mathbb{Z}$, the set of integers. Then the usual relation "less than or equal to" denoted by $\leq$, defined as

$$\leq = \{(x, y) : x, y \in \mathbb{Z}, \text{ and } x \text{ is less than or equal to } y\}$$

is a binary relation on $\mathbb{Z}$.

2. Let A be the set of living persons on earth. The relation "is son of" given by

$$\text{is son of } = \{(x, y) : x, y \in A \text{ and } x \text{ is a son of } y\}$$

is a binary relation on A.

3. Let $\mathbb{P}(X)$ be the power set of a set X. Then the relation "is a subset of" denoted by $\subseteq$, given by

$$\subseteq = \{(A, B) : A, B \in \mathbb{P}(X), \text{ and } A \text{ is a subset of } B\}$$

is a binary relation on $\mathbb{P}(X)$.

4. The relation "is successor of", denoted by S^+, given by

$$S^+ = \{(x, y) : x \text{ and } y \text{ are positive integers and } x \text{ is a successor of } y\},$$

where the successor of n is $n + 1$ for every positive integer n, is a binary relation on the set of positive integers. The ordered pair (1,1) is not a member of S^+ as 1 is not a successor of 1.

There are certain relations in which mathematicians are usually more interested. One such class of relations is the class of equivalence relations.

Definition 1.4.3 A relation R on a set X is called

(i) reflexive if $(x, x) \in R$ for all x in X.

(ii) symmetric if $(x, y) \in R$ implies that $(y, x) \in R$ for all x, y in X.

(iii) transitive if $(x, y) \in R$ and $(y, z) \in R$ implies that $(x, z) \in R$ for all x, y, z in X.

Definition 1.4.4 A relation is called an equivalence relation if it is reflexive, symmetric and transitive.

Examples 1.4.5

1. The relation of equality "=" on any set is an equivalence relation. Since $x = x$, for all x, the relation $=$ is reflexive. Since $x = y$ implies that $y = x$, the relation $=$ is symmetric. Since $x = y$ and $y = z$ implies that $x = z$, the relation $=$ is transitive.

2. The relation $\subseteq$ in Example 1.4.2 (3) is reflexive and transitive, but not symmetric.

3. The usual relation "less than or equal to" on the set of real numbers $\mathbb{R}$ is reflexive and transitive but not symmetric. So it is not an equivalence relation. The relation "less than" on the set of real numbers is not reflexive, but it is a transitive relation.

An equivalence relation on a set X partitions X into disjoint sets. To see this, assume that "tilde", denoted by $\sim$, is an equivalence relation on a set X, and x is an element of X. Then the set

$$E = \{y \in X : y \sim x\}$$

is called an equivalence class determined by x. E is non-empty since x is in E. The equivalence classes have the following property.

Lemma 1.4.6 Any two equivalence classes are either disjoint or equal.

Proof. Let $\sim$ be an equivalence relation on a set X. Let E and E' be two equivalence classes. Suppose that E and E' are not disjoint. Then there is an element y in $E \cap E'$. Let z be any element in E'. Then $y \sim z$. Since y is in E, z is in E. So that E' is contained in E. Similarly, E is contained in E'. Therefore, E and E' are equal.

Since each element x is in some equivalence class, for instance, x is in the equivalence class determined by x, it follows that an equivalence relation on X partitions X into disjoint equivalence classes.

Definition 1.4.7 A partition of a set X is a set $\mathcal{A}$ of disjoint subsets of X such that

$$X = \bigcup_{A \in \mathcal{A}} A.$$

As already noted, an equivalence relation on a set X determines a partition of X. On the other hand, given a partition of a set X, there is a unique equivalence relation whose equivalence classes are just the sets in the partition. Let $\mathcal{A}$ be a partition of X. We define a relation R on X by setting xRy if x and y are in a set in $\mathcal{A}$. It is easily seen that R is reflexive and symmetric. Now assume that xRy and yRz. Then there exist sets A_1 and A_2 in $\mathcal{A}$ such that $x, y \in A_1$ and $y, z \in A_2$. Since y is common to A_1 and A_2, it follows that $A_1 = A_2$. Therefore, xRz. Thus, R is an equivalence relation on X whose equivalence classes are precisely the elements of the partition $\mathcal{A}$. To show uniqueness, assume that R_1 and R_2 are two equivalence relations giving the partition $\mathcal{A}$. Let x be a point in X. Then the equivalence class determined by x with respect to R_1 is some set $A_1 \in \mathcal{A}$. The set A_1 contains x. Also the equivalence class of x with respect to R_2 is a set A_2 in $\mathcal{A}$. Since A_2 also contains x, it follows that $A_1 = A_2$. This means that for any y in X yR_1x if and only if yR_2x. Since x is arbitrary, $R_1 = R_2$.

Example 1.4.8 Let $\mathbf{I}$ be the set of all integers. Let m be a fixed positive integer. We define a relation $\sim$ on $\mathbf{I}$ by

$$a \sim b \text{ if } a - b \text{ is an integral multiple of } m.$$

(Here the terminology and symbols have their usual meanings found in arithmetic of integers.) It is easily seen that the relation $\sim$ is an equivalence relation. Denoting the equivalence class determined by x with respect to $\sim$ by $[x]$, the equivalence classes are

$$\begin{aligned}
[0] &= \{0, \pm m, \pm 2m, \ldots\} \\
[1] &= \{1, \pm m + 1, \pm 2m + 1, \ldots\} \\
[2] &= \{2, \pm m + 2, \pm 2m + 2, \ldots\} \\
&\vdots \\
[m-1] &= \{m - 1, \pm m + m - 1, \pm 2m + m - 1, \ldots\}.
\end{aligned}$$

There are m distinct equivalence classes. The relation $\sim$ in this example is usually called "congruent modulo m", and symbolically written as $a \equiv b \pmod{m}$.

Example 1.4.9 In example 1.4.5. (1), the relation of equality $=$ is shown to be an equivalence relation on any set A. The equivalence class of any element x in A is the singleton $\{x\}$. Thus, there are as many equivalence classes for this relation as there are

points in A. On the other hand, the relation "$x \sim y$ for all points x, y in A" is also an equivalence relation on A. There is only one equivalence class for all elements with respect to this relation which is equal to A. The above two equivalence relations are two extremes, and every other equivalence relation lies between these two extremes.

Example 1.4.10 The set $\mathcal{A}$ of all circles in the plane is not a partition of the plane. The two circles in $\mathcal{A}$ may intersect without being identical. However, the set of all lines parallel to a given line in the plane is a partition of the plane.

1.5 Mappings

Mappings are the rules of associations among the objects: Given any two set A and B, a rule is given that relates a unique element of B to each element in A. Thus, a mapping is a special kind of a relation. Formally, we have the following definition.

Definition 1.5.1 Let A and B be two sets. A mapping f from A to B is a subset of $A \times B$ satisfying the following two conditions:

1. For each point x in A, there exists a point y in B such that $(x, y) \in f$.
2. For all (x_1, y_1) and (x_2, y_2) in f, $x_1 = x_2$ implies that $y_1 = y_2$.

The set A in the above definition is called the domain of f, and is denoted by $\text{dom}(f)$, and the set B is called the codomain of f, and is denoted by $\text{cod}(f)$. The range of f, denoted by $\text{range}(f)$, is defined as

$$\text{range}(f) = \{y \in B : \text{ there exists a point } x \text{ in } A \text{ such that } (x, y) \in f\}.$$

Obviously, $\text{range}(f) \subseteq \text{cod}(f)$. When (x, y) is in f, we say that y is the image of x under f, and write $y = f(x)$. Also x is a preimage of y under f. This is entirely possible that a point y in B has none, exactly one, or several preimages. It may be noticed however that there is some shift in terminology while moving from the topic of relations to that of mappings; when $y = f(x)$, then instead of saying that y is related to x by f, the emphasis is placed on how the image y is obtained from x.

The mapping f may be written equivalently as

$$f = \{(x, f(x)) : x \in A\}.$$

The set f is often called the graph of f. Thus, a mapping and the graph of the mapping represent the same set of points of $A \times B$.

Instead of writing a mapping f as a set of ordered pairs, we usually write $f : A \to B$, and read it as: f is a mapping from A to B, and give the rule of forming images. Thus, if $(x, y) \in f$, then $y = f(x)$ is the corresponding rule of forming the image y in B of the point x in A. For example, let $f : \mathbb{R} \to \mathbb{R}$ be the mapping defined by the rule $y = f(x) = \sin x$. Here, the number $\sin x$ is the image of x under f. Equivalently, the same mapping f may be denoted in the form of ordered pairs by

$$f = \{(x, \sin x) : x \in \mathbb{R}\}.$$

As an another example, the mapping $f : \mathbb{R} \to \mathbb{R}$ defined by the rule $f(x) = 3x^2 + 2x + 3$ for all x in $\mathbb{R}$, in ordered pair form can be written equivalently as

$$f = \{(x, 3x^2 + 2x + 3) : x \in \mathbb{R}\}.$$

Usually, the word 'function' is used for a mapping when the domain and codomain sets are subsets of real or complex numbers.

Example 1.5.2 In order to further illustrate the concept of a mapping, define $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$. Consider the following subsets of $A \times B$.

$$f_1 = \{(a, 1), (b, 2)\},$$
$$f_2 = \{(a, 1), (b, 1), (c, 2)\},$$
$$f_3 = \{(a, 1), (a, 2), (b, 1), (c, 3)\},$$
$$f_4 = \emptyset.$$

The set f_1 is not a mapping since it violates the condition (1) in the definition of a mapping; there is no ordered pair corresponding to c in A. The set f_2 represents a mapping. The set f_3 is not a mapping since $(a, 1)$ and $(a, 2)$ are in f_3, and the first co-ordinates in these two ordered pairs are same, but the second co-ordinates are different, thus violating the conditions (2) of a mapping. The set f_4 is again not a mapping; it violates the condition (1).

Definition 1.5.3 Let $f : A \to B$ and $g : A \to B$ be two mappings. Then f and g are identical mappings, written $f = g$, if $f(x) = g(x)$ for all x in A.

There are several ways of generating new mappings from the given ones. One of these is to form composition of mappings.

Definition 1.5.4 Let $f : A \to B$ and $g : B \to C$ be two mappings. Then the composition of f and g, denoted $g \circ f$, is a mapping from A to C, that is, $g \circ f : A \to C$, defined by

$$(g \circ f)(x) = g(f(x)) \text{ for all } x \text{ in } A.$$

Equivalently, in terms of ordered pairs, $g \circ f$ is a subset of $A \times C$ given by

$$g \circ f = \{(x, z) : \text{ there exists a point } y \text{ in } B \text{ such that } (x, y) \in f \text{ and } (y, z) \in g\}.$$

As an illustration, let us consider the mappings $f : \mathbb{N} \to \mathbb{N}$ defined by $f(x) = x^2$, and $g : \mathbb{N} \to S$, where S is the set of all circles in the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ defined by $g(x) = C_x$, where C_x is a circle in S with centre at origin and radius x. The composition $g \circ f$ of f and g is the mapping from $\mathbb{N}$ to S given by

$$(g \circ f)(x) = g(f(x)) = g(x^2) = C_{x^2}$$

which is a circle in S with centre at the origin and radius x^2.

The second way of generating a new mapping from a given mapping is to restrict the later on a subset of its domain. More specifically, we have the following.

Definition 1.5.5 Let $f : A \to B$ be a mapping, and A_0 be a subset of A. Then the restriction of f to A_0 is the mapping $g : A_0 \to B$ such that $g(x) = f(x)$ for each x in A_0. That is

$$g = \{(x, f(x)) : x \in A_0\}.$$

The restriction of f to A_0 is denoted by f_{A_0} or $f|_{A_0}$, that is, $f_{A_0} = g$.

Example 1.5.6 Define $A = \mathbb{R} - \{(2n + 1)\frac{\pi}{2} : n \text{ is an integer}\}$. Let $f : A \to \mathbb{R}$ be the mapping given by $f(x) = \tan x$. Assume that $A_0 =]0, \frac{\pi}{2}[$. Let $g : A_0 \to \mathbb{R}$ be the mapping

defined by $g(x) = \tan x$ for all x in $]0, \frac{\pi}{2}[$. Since $f(x) = g(x)$ for all x in A_0, g is the restriction of f to A_0.

Definition 1.5.7 Let $f : A \to B$ be a mapping. If A_0 is a subset of A, then the image of A_0 under f, denoted by $f(A_0)$, is defined as the set of points in B which are images of the points in A_0 under f, that is

$$f(A_0) = \{y \in B : \text{ there exists a point } x \text{ in } A_0 \text{ such that } f(x) = y\}.$$

If B_0 is a subset of B, then the preimage of B_0, denoted by $f^{-1}(B_0)$, is the set of points in A which are mapped by f in B_0, that is

$$f^{-1}(B_0) = \{x \in A : f(x) \in B_0\}.$$

A mapping $f : A \to B$ is said to be injective or one-one if

$$f(x) = f(y) \text{ implies that } x = y.$$

A mapping $f : A \to B$ is said to be surjective or onto if for each point y in B, there exists an x in A such that $f(x) = y$.

A mapping $f : A \to B$ is said to be bijective if it is both injective and surjective. A bijective mapping is also called a one-to-one correspondence.

If $f : A \to B$ is a bijective mapping, we define a mapping $f^{-1} : B \to A$ by $f^{-1}(y) = x$ if $f(x) = y$. The mapping f^{-1} is called the inverse mapping of f.

Examples 1.5.8

1. The mapping $f :]-\frac{\pi}{2}, \frac{\pi}{2}[\to \mathbb{R}$ defined by $f(x) = \tan x$ is both injective and surjective.
2. The mapping $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin x$ is neither injective nor surjective.
3. The mapping $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = e^x$ is injective, but not surjective.
4. For any set A, the mapping $f : \mathbb{P}(A) \to \mathbb{P}(A)$ defined by $f(B) = A - B$ for all B in $\mathbb{P}(A)$ is bijective.
5. The mapping $f : \mathbb{R} \to [-1, 1]$ defined by $f(x) = \sin x$ is surjective, but not injective.
6. If $f : A \to B$ and $g : B \to C$ are mappings such that $g \circ f$ is surjective, then g is surjective, but f need not be surjective. For example, define $A = \{a, b\}$, $B = \{1, 2\}$, $C = \{h\}$ and let f and g be mappings such that $f(a) = f(b)$, and $g(1) = g(2) = h$. Then $g \circ f$ is surjective while f is not.
7. If $f : A \to B$ and $g : B \to C$ are mappings such that $g \circ f$ is injective, then f is injective but g may not be injective. For example, define $A = \{a, b\}$, $B = \{1, 2, 3\}$, $C = \{x, y\}$, $f : A \to B$ by $f(a) = 1$, $f(b) = 2$, $g : B \to C$ by $g(1) = x$, $g(2) = y$, $g(3) = x$.

There are certain mappings which occur so frequently that they are given special names. A mapping $i : A \to A$ defined by $i(x) = x$ for all x in A is called an identity mapping. There is, obviously, only one such mapping f on each set A. The graph of the identity mapping $i : A \to A$ is the set $\{(x, x) : x \in A\}$ which is known as diagonal of $A \times A$.

A mapping $f : A \to B$ is called a constant mapping if there exists an element y in B such that $f(x) = y$ for all x in A.

For any sets A and B, the mappings $\pi_1 : A \times B \to A$ defined by $\pi_1(x, y) = x$, and $\pi_2 : A \times B \to B$ defined by $\pi_2(x, y) = y$ are called projection mappings or projections; π_1 is called the first projection and π_2 the second projection. The projections are, obviously, surjective.

Let A and B be two sets such that $A \subseteq B$. Then the mapping $i : A \to B$ defined by $i(x) = x$ for all x in A is called the inclusion mapping.

Let A and B be sets, and A_0 be a subset of A. Let $f : A \to B$ and $g : A_0 \to B$ be mappings. Then f is called an extension of g *if* $f(x) = g(x)$ for all x in A_0. In general, there may be several extensions of g.

Let A be a subset of a set X. Then the mapping $\chi_A : X \to \{0, 1\}$ defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A \end{cases}$$

is called the characteristic mapping associated with A.

Theorem 1.5.9 Let $f : A \to B$ and $g : B \to A$ satisfy $g \circ f = i_A$, where i_A is the identity mapping defined on A. Then f is injective, and g is surjective.

Proof. $f(x) = f(x')$ implies that $g(f(x)) = g(f(x'))$. Therefore, $x = (g \circ f)(x) = (g \circ f)(x') = x'$. g is surjective since for any x in A, $g(f(x)) = x$, and $f(x) \in B$.

Theorem 1.5.10 Let $f : A \to B$ be a mapping. If there are mappings $g : B \to A$ and $h : B \to A$ such that $g \circ f = i_A$ and $f \circ h = i_B$. Then f is bijective, and $g = h = f^{-1}$.

Proof. The proof is left as an exercise.

Definition 1.5.11 Let $\mathcal{A}$ be a family of sets. An indexing mapping for $\mathcal{A}$ is a surjection f from some set J, called the index set, to $\mathcal{A}$. The family $\mathcal{A}$ together with the indexing mapping f is called an indexed family of sets.

Given an element i in the index set J, we denote $f(i)$ in $\mathcal{A}$ by A_i. Then the indexed family itself is denoted by the symbols

$$\{A_i\}_{i \in J}$$

which is read as "the family of all A_i's as i ranges over the indexed set J". Sometimes, we simply write $\{A_i\}$ when the underlying index set is understood.

In the definition of indexing mapping, f need not be injective, it is entirely possible that $A_i = A_j$ for distinct i and j in J.

Now we define the union and the intersection of indexed family of sets. Let $\{A_i\}_{i \in J}$ be an indexed family of sets. The union of all sets in this family, denoted by $\bigcup_{i \in j} A_i$, is defined as

$$\bigcup_{i \in J} A_i = \{x : x \in A_i \text{ for some } i \text{ in } J\}.$$

The intersection of all sets in the family $\{A_i\}_{i \in J}$, denoted by $\bigcap_{i \in J} A_i$, is defined as

$$\bigcap_{i \in J} A_i = \{x : x \in A_i \text{ for all } i \text{ in } J\}.$$

Examples 1.5.12

1. Let $\mathbf{J} = \mathbb{Z}_+$, where $\mathbb{Z}_+$ is the set of positive integers. For each $n \in \mathbb{Z}_+$, define $A_n =]-n, n[$. Then

$$\bigcup_{n\in\mathbb{Z}_+} A_n = \mathbb{R}, \quad \text{and} \quad \bigcap_{n\in\mathbb{Z}_+} A_n =]-1, 1[.$$

   It is a customary to denote $\bigcup_{n\in\mathbb{Z}_+} A_n$ by $\bigcup_{n=1}^{\infty} A_n$, and $\bigcap_{n\in\mathbb{Z}_+} A_n$ by $\bigcap_{n=1}^{\infty} A_n$.

2. We define $A_a = \{a\}$ for each $a \in \mathbb{R}$. Then $\bigcup_{a\in\mathbb{R}} A_a = \mathbb{R}$. Here $\mathbb{R}$ is also acting as an index set.

3. If J is the empty set, then the family $\{A_i\}_{i\in\mathbf{J}}$ of sets is an empty family. Its union $\bigcup_{i\in\mathbf{J}} A_i$ is the empty set.

4. The intersection of the family defined in (2), $\bigcap_{a\in\mathbb{R}} A_a$ is the empty set.

5. Let $\{A_i\}_{i\in\mathbf{J}}$ be a family of sets. If J is empty, then the family $\{A_i\}_{i\in J}$ is an empty family of sets. Which set is equal to $\bigcap_{i\in\mathbf{J}} A_i$, where J is empty? Note that the similar problem we encountered in defining arbitrary intersection. In the case when the family $\{A_i\}_{i\in\mathbf{J}}$ consists of subsets of a set X, we have the elements of X at our disposal to be considered for the membership in the intersection $\bigcap_{i\in\mathbf{J}} A_i$. Thus, when J is empty, every element of X qualifies to be the member of the intersection $\bigcap_{i\in\mathbf{J}} A_i$ since there is no A_i in the family $\{A_i\}_{i\in\mathbf{J}}$ to which it does not belong. Therefore,

$$\bigcap_{i\in\mathbf{J}} = X.$$

The concept of indexed family of sets becomes indispensable when we extend the notion of Cartesian product to arbitrary families of sets.

Let J be an index set. For any set X, we define a J-tuple to be a mapping $x : J \to X$. If for each i, $x(i)$ is denoted by x_i, then J-tuple x itself is denoted by $\langle x_i\rangle_{i\in J}$. The set of all J-tuples in X is denoted by $X^{\mathbf{J}}$.

Definition 1.5.13 Let $\{A_i\}_{i\in\mathbf{J}}$ be an indexed family of sets, and $X = \bigcup_{i\in\mathbf{J}} A_i$. The Cartesian product or simply the product of the family $\{A_i\}_{i\in\mathbf{J}}$, denoted by $\prod_{i\in\mathbf{J}} A_i$, is the set of all J-tuples $\langle x_i\rangle_{i\in\mathbf{J}}$ in X such that $x_i \in A_i$ for all $i \in J$.

Since each J-tuple $\langle x_i\rangle_{i\in\mathbf{J}}$ corresponds to the mapping $x : J \to X$ defined by $x(i) = x_i$, equivalently, the Cartesian product $\prod_{i\in J} A_i$ is the set of all mappings $x : J \to X$ such that $x(i) \in A_i$ for all $i \in J$.

If $A_i = X$ for all $i \in J$, then we write $\prod_{i\in\mathbf{J}} A_i$ as $X^{\mathbf{J}}$. If $J = \mathbb{Z}_+$, then the Cartesian product of the family $\{A_i\}_{i\in\mathbb{Z}_+}$ is usually denoted by $\prod_{i=1}^{\infty} A_i$. Similarly, if $J = \{1, 2, \ldots, n\}$, then the notation for the Cartesian product is $\prod_{i=1}^{n} A_i$ or $A_1 \times A_2 \times \ldots \times A_n$. If $\mathbf{J} = \mathbb{Z}_+$, $\mathbb{Z}_+$-tuple is usually called an ω-tuple, and $X^{\mathbb{Z}_+}$ is written as X^{ω}. An ω-tuple is also called a sequence. If $J = \{1, 2, \ldots, n\}$, then a $\mathbf{J}$-tuple is also called a n-tuple. When $\mathbf{J}$ is understood from the context, we just write $\prod A_i$ for the product. Also, when $\mathbf{J} = \{1, 2, \ldots, n\}$ and $A_i = A$ for all $i = 1, 2, \ldots, n$, then $\prod_{i=1}^{n} A_i$ is usually denoted by A^n.

An operation over a set is also a mapping in which we tell not only that an element is the image of some element, we also tell how to compute that image. For example, the operation of addition $+$ on the set of natural number $\mathbb{N}$ is the mapping: $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ which maps an ordered pair (a, b) in $\mathbb{N} \times \mathbb{N}$ to the element $a + b$ in $\mathbb{N}$, and this mapping $+$ also tells us how to compute $a + b$; for instance, the image 5 of the element $(2, 3)$ in $\mathbb{N} \times \mathbb{N}$. Such operations provide an internal algebraic structure to the underlying sets. These algebraic structures occur in several places in modern mathematics. For their own sake also these algebraic structures are investigated extensively under the topic "Modern Algebra". Here we cover some elementary definitions which are relevant to our purpose.

Definition 1.5.14 F be a non-empty set. Let $+ : F \times F \to F$ and $\cdot : F \times F \to F$ be two mappings called addition and multiplication operations, respectively (we agree to write $+(a, b) = a + b$ and $\cdot(a, b) = ab$) such that for all a, b and c in F

(i) $(a + b) + c = a + (b + c)$,

(ii) there is an element 0 in F such that $a + 0 = 0 + a = a$, (0 is called the zero element or the additive identity),

(iii) there is an element $-a$ in F such that $a + (-a) = (-a) + a = 0$,

(iv) $a + b = b + a$,

(v) $a(bc) = (ab)c$,

(vi) $(a + b)c = ac + bc$,
$a(b + c) = ab + ac$.

Then the triad $(F, +, \cdot)$ is called a ring.

1. A ring $(F, +, \cdot)$ is called a ring with unit element 1 in F if
   $a1 = 1a = a$ for all a in F.

2. A ring $(F, +, \cdot)$ is called a commutative ring if
   $ab = ba$ for all a and b in F.

3. A commutative ring $(F, +, \cdot)$ is called an integral domain if
   $ab = 0$ implies that $a = 0$ or $b = 0$ for all a and b in F.

4. A commutative ring with unit element $(F, +, \cdot)$ with the property that for each non-zero element a in F, there exists an element a^{-1} in F such that
   $$aa^{-1} = a^{-1}a = 1$$
   is called a field.

For example, the set of integers $\mathbb{Z}$ with usual addition $+$ and multiplication $\cdot$ is an integral domain which is not a field.

$(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$ are the examples of fields, where $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are sets of rational, real and complex numbers, respectively, and $+$ and $\cdot$ are usual addition and multiplication operations on them.

A subset H of a field F is said to be subfield of F if under the restriction of the operations in F, H itself is a field.

EXERCISE

1. Let $A, B, C, \ldots$ be subsets of a set X. Prove the following:
   (a) $A - A = \emptyset$, $A - \emptyset = A$.
   (b) $A \cup A = A$, $A \cap A = A$.
   (c) $A \cup B = B \cup A$, $A \cap B = B \cap A$.
   (d) The relations $A \subseteq B$, $A \cup B = B$ and $A \cap B = A$ are same.
   (e) $A \subseteq A \cup B$, $A \cap B \subseteq A$.
   (f) The two relations $A \subseteq C$ and $B \subseteq C$ are equivalent to the relation $A \cup B \subseteq C$.
   (g) $A \cup (B \cup C) = (A \cup B) \cup C$, written as $A \cup B \cup C$,
   $A \cap (B \cap C) = (A \cap B) \cap C$, written as $A \cap B \cap C$.
   (h) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
   (i) $(A')' = A$.
   (j) $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$.
   (k) $A \subseteq B$ if and only if $A' \supseteq B'$.
   (l) $A \cap B = \emptyset$ if and only if $A \subseteq B'$ if and only if $B \subseteq A'$.
   (m) $A \cup B = X$ if and only if $A' \subseteq B$ if and only if $B' \subseteq A$.

2. Let $\{A_i\}_{i \in J}$ be a family of subsets of a set X, and $\{B_i\}_{i \in K}$ be a family of subsets of a set Y. Prove the following identities:
   (a) $\left(\bigcup_{i \in J} A_i\right)' = \bigcap_{i \in J} A_i'$, $\left(\bigcap_{i \in J} A_i\right)' = \bigcup_{i \in J} A_i'$.
   (b) $\left(\bigcap_{i \in J} A_i\right) \bigcup \left(\bigcap_{j \in K} B_j\right) = \bigcap_{(i,j) \in J \times K} \left(A_i \bigcup B_j\right)$.
   (c) $\left(\bigcup_{i \in J} A_i\right) \bigcap \left(\bigcup_{j \in K} B_j\right) = \bigcup_{(i,j) \in J \times K} \left(A_i \bigcap B_j\right)$.

3. (i) The symmetric difference of two sets A and B, denoted by $A \Delta B$, is defined as
   $$A \Delta B = (A - B) \cup (B - A).$$
   A ring of sets is non-empty class $\mathcal{A}$ of sets such that if A and B are elements in $\mathcal{A}$, then $A \Delta B$ and $A \cap B$ are also in $\mathcal{A}$. Show that a ring $\mathcal{A}$ contains the empty set, $A \cup B$ and $A - B$.

(ii) If $\mathcal{A}$ is a non-empty class of sets such that for any sets A and B in $\mathcal{A}$, $A \cup B$ and $A - B$ are also in $\mathcal{A}$, then $\mathcal{A}$ is a ring of sets.

(iii) Show that the class of all finite subsets (including the empty set) of an infinite set is a ring of sets.

(iv) Show that the class of all finite unions of closed open intervals on the real line is a ring of sets.

4. Let $A, B, C, \ldots$ be sets. Prove the following.

   (a) $(A - C) \cap (B - C) = (A \cap B) - C$.

   (b) If $A \subseteq B$, then $A \times C \subseteq B \times C$.

   (c) $A \times (B \cup C) = (A \times B) \cup (A \times C)$.
   $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

   (d) $(A \times B) \cap (S \times T) = (A \cap S) \times (B \cap T)$.

   (e) $A - B = A$ if and only if $A \cap B = \emptyset$.

   (f) $(A \times C) \cup (B \times C) = (A \cup B) \times C$.

5. If A contains n elements, then the power set of A, $\mathbb{P}(A)$, contents 2^n elements. Prove this statement using induction.

6. The statement $A = \emptyset$ is equivalent to the statement "there is no x such that $x \in A$". Rewrite each of the following without using the symbol $\emptyset$.

   (a) $A \neq \emptyset$.

   (b) $A \cap B = \emptyset$.

   (c) If $A \neq \emptyset$, and $B \neq \emptyset$, then $A \cup B \neq \emptyset$.

7. Find the elements of $\mathbb{P}(\mathbb{P}(\mathbb{P}(\mathbb{P}(\emptyset))))$.

8. Define $(x, y, z) = ((x, y), z)$. Now prove the following statement. $(x, y, z) = (a, b, c)$ if and only if $x = a$, $y = b$, $z = c$ for any sets x, y, z, a, b and c.

9. Let X be a set. Let $g : X \to X$ be a mapping, and $a \in X$. Show that there is a mapping $\emptyset$ with domain $\mathbb{N}_0$ such that $\phi(n) = g^n(a)$, where g^n denotes the composition of g with itself n-times for $n > 0$, and $g^0(a) = a$.

10. Let S be a relation from X to Y, and R be a relation from Y to Z. The composition of R and S, denoted $R \circ S$, is a relation from X to Z defined by

$$R \circ S = \{(x, z) \in X \times Z : \text{ there is a } y \in Y \text{ such that } (x, y) \in S \text{ and } (y, z) \in R\}.$$

For any subset A of X, the image of A under S, denoted by $S(A)$, is defined by

$$S(A) = \{y \in Y : \text{ there is a } x \text{ in } A \text{ such that } (x, y) \in S\}.$$

For any subset B of Y, the pre-image of B under S, denoted by $S^{-1}(B)$, is defined as

$$S^{-1}(B) = \{x \in X : \text{ there is a } y \text{ in } B \text{ such that } (x, y) \in S\}.$$

(a) Prove that for any subset A of X, $(R \circ S)(A) = R(S(A))$. Similarly, for any subset B of Z, $(R \circ S)^{-1}(B) = S^{-1}(R^{-1}(B))$.

(b) If $f : X \to Y$ and $g : Y \to Z$ are both mappings, then show that $g \circ f$ is a mapping from X into Z.

11. Let X and Y be non-empty sets, and R be a relation from X to Y. The inverse relation R^{-1} of R is a relation from Y to X defined by

$$R^{-1} = \{(y, x) : (x, y) \in R\}.$$

(a) Show that $R = R^{-1}$ if and only if $X = Y$ and R is a symmetric relation.

(b) If $f : X \to Y$ is a mapping, then f^{-1} is a mapping of the range B of f onto X if and only if f is injective. Verify that if f is bijective, then $f^{-1} \circ f$ and $f \circ f^{-1}$ are identity mappings on X and Y, respectively.

(c) Let $f : X \to Y$ be a mapping. Show that there exists a mapping $g : Y \to X$ such that $g \circ f = i_X$ if and only if f is injective. Prove also that there exists a mapping $h : Y \to X$ such that $f \circ h = i_Y$ if and only if f is surjective. Conclude from this that if there exist mappings g and h from Y into X such that

$$g \circ f = i_X \quad \text{and} \quad f \circ h = i_Y,$$

then f is bijective, and $g = h = f^{-1}$.

12. Let $f : A \to B$ be a mapping, and $A_i \subseteq A$, $B_j \subseteq B$ for i in an index set Λ_1 and j in an index set Λ_2. Prove the following.

(a) $A_i \subseteq A_j$ implies that $f(A_i) \subseteq f(A_j)$.

(b) $f(A_i \cup A_j) = f(A_i) \cup f(A_j)$, $f\left(\bigcup_{i \in \Lambda_1} A_i\right) = \bigcup_{i \in \Lambda_1} f(A_i)$.

(c) $f(A_i \cap A_j) \subseteq f(A_i) \cap f(A_j)$, $f\left(\bigcap_{i \in \Lambda_1} A_i\right) \subseteq \bigcap_{i \in \Lambda_1} f(A_i)$. Give an example where equality fails.

(d) $f(A_i - A_j) \supseteq f(A_i) - f(A_j)$. Give an example where equality fails.

(e) $B_i \subseteq B_j$ implies that $f^{-1}(B_i) \subseteq f^{-1}(B_j)$.

(f) $f^{-1}(B_i \cup B_j) = f^{-1}(B_i) \cup f^{-1}(B_j)$, $f^{-1}\left(\bigcup_{i \in \Lambda_2} B_i\right) = \bigcup_{i \in \Lambda_2} f^{-1}(B_i)$.

(g) $f^{-1}(B_i \cap B_j) = f^{-1}(B_i) \cap f^{-1}(B_j)$, $f^{-1}\left(\bigcap_{i \in \Lambda_2} B_i\right) = \bigcap_{i \in \Lambda_2} f^{-1}(B_i)$.

(h) $f^{-1}(B_i - B_j) = f^{-1}(B_i) - f^{-1}(B_j)$.

(i) $A_i \subseteq f^{-1}(f(A_i))$, and that equality holds if f is injective.

(j) $f(f^{-1}(B_i)) \subseteq B_i$, and that equality holds if f is surjective.

13. Let $f : A \to B$ and $g : B \to C$ be mappings. Prove the following.

(a) If $C_0 \subseteq C$, then $(g \circ f)^{-1}(C_0) = f^{-1}(g^{-1}(C_0))$.

(b) If f and g are injective, then $g \circ f$ is injective.

(c) If f and g are surjective, then $g \circ f$ is surjective.

14. Let $f : \mathbb{R} \to \mathbb{R}$ be the mapping $f(x) = x^3 - x$. By restricting the domain and codomain, obtain from f a bijective mapping g. Show that there are several possible choices for g.

15. Let $f : X \to Y$ be a mapping. Show that the following statements are equivalent.

(a) f is injective.

(b) For any subset A of X, $f^{-1}(f(A)) = A$.

(c) For any pair of subsets A, B of X, $f(A \cap B) = f(A) \cap f(B)$.

(d) For any pair of subsets A, B of X such that $A \cap B = \emptyset$, $f(A) \cap f(B) = \emptyset$.

(e) For any pair of subsets A, B of X such that $B \subseteq A$, $f(A - B) = f(A) - f(B)$.

16. Let $f : X \to Y$ and $g : Y \to Z$ be two bijections. Then $g \circ f : X \to Z$ is a bijection, and that the inverse mapping $(g \circ f)^{-1} : Z \to X$ is the same as the composition $f^{-1} \circ g^{-1} : Z \to X$, that is, $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

17. Let $f : A \to B$ be a mapping. Show that the mapping $g : A \to A \times B$ defined by $g(x) = (x, f(x))$ is injective.

18. Let $f : A \to B$, $g : B \to C$, $h : C \to D$ be mappings such that $g \circ f$ and $h \circ g$ are bijective. Show that f, g and h are all bijective mappings.

19. Let $f : A \to B$, $g : B \to C$ and $h : C \to A$ be mappings. Show that if among $h \circ g \circ f$, $g \circ f \circ h$, $f \circ h \circ g$ two are surjective and the third is injective, or two are injective and the third is surjective, then all the three mappings f, g and h are bijective.

20. (a) Let $f : A \to B$ be a mapping. Show that the relation on A defined by $x \sim x'$ if $f(x) = f(x')$ is an equivalence relation. The equivalence classes of this relation are called level sets of f.

(b) Let $\sim$ be an equivalence relation on a set A, and $P_\sim = \{[x] : x \in A\}$ be the set of equivalence classes determined by the relation $\sim$. $P_\sim$ is sometimes called the quotient space of A modulo $\sim$. The mapping $\pi : A \to P_\sim$ defined by $\pi(x) = [x]$ is called the natural projection of A onto $P_\sim$. Show that the set of level sets of π coincides with $P_\sim$.

21. Give an example of a relation which is (a) reflexive but not symmetric or transitive (b) symmetric but not reflexive or transitive (c) transitive but not reflexive or symmetric (d) reflexive and symmetric but not transitive (e) reflexive and transitive but not symmetric (f) symmetric and transitive but not reflexive.

22. If $f : A \to B$ be a mapping, and χ_E be the characteristic mapping of $E \subseteq B$. Then $\chi_E \circ f$ is the characteristic mapping of which subset of A?

23. The sets $A_1 \times A_2 \times A_3$, $(A_1 \times A_2) \times A_3$ and $A_1 \times (A_2 \times A_3)$ are distinct sets. Set up a bijective correspondence between any two of them.

24. Let U be the set $\{1, 2, \ldots, n\}$ for any arbitrary positive integer n. How many subsets of U are there. How many relations can be defined of the form $A \subseteq B$ on $\mathbb{P}(U)$?

25. A bijective mapping from a finite set onto itself is called a permutation of the set. Write down all the permutations of the set $S_4 = \{1, 2, 3, 4\}$.

26. Spot the flaw in the following arguments: Let R be transitive and symmetric. Then xRy and yRz implies that xRz for all x, y and z. Also xRy implies that yRx holds for all x and y. Now take any x and y such that xRy, then by the preceding argument xRx. Hence, R is reflexive.

27. Suppose that G and H are equivalence relations on A, and that $G \cup H$ is also an equivalence relation on A. Prove that for each $x \in A$, $[x]_{G\cup H} = [x]_G \cup [x]_H$, where $[x]_G$ is the equivalence class of x with respect to G.

28. Let G and H are equivalence relations on A. Prove that for each $x \in A$, $[x]_{G\cap H} = [x]_G \cap [x]_H$.

29. Let $\{R_i\}_{i\in I}$ be a family of equivalence relations on A. Prove that $\bigcap_{i\in I} R_i$ is an equivalence relation on A. Check whether $\bigcup_{i\in I} R_i$ is also an equivalence relation on A.

30. Let L be the set of all straight lines in the plane. Let G and H be the following relations on L.

    $G = \{(l, m) \in L^2 : l \text{ is parallel to } m\}$,

    $H = \{(l, m) \in L^2 : l \text{ is perpendicular to } m\}$.

    Prove the following.

    (a) G is an equivalence relation on L.

    (b) $H \circ G = H$, and $G \circ H = H$.

    (c) $G \cup H$ is an equivalence relation; describe its equivalence classes.

31. Show that the equation

$$f(m, n) = 2^m(2n + 1) - 1$$

    defines a one-to-one correspondence between $\mathbb{N}_0 \times \mathbb{N}_0$ and $\mathbb{N}_0$.

1.6 Predicates, Quantifiers, Negations, Converse and Contrapositive

An experienced reader of mathematics would realise that the way mathematical expressions are formed is, infact, same in all branches of mathematics. Even the rules of inference that are used to derive new expressions from the given ones are the same. These and others are concerns we deal in mathematical logic, and this is not a place to discuss these constructs of mathematical logic in any detail, but some elementary bare minimum introduction of the

terminology of the area useful to our present purpose is, perhaps, inescapable in order to have a unified view.

In (predicate) logic, an assertive sentence is usually called a statement. For instance, the following sentences are statements.

1. All men are mortal.
2. 20 is not less than 30.
3. All squared real numbers are positive.
4. Some integers are less than ten.
5. If Ramesh is in Delhi, then he will appear in the examination.

Each statement is assigned a truth value 'True' or T, or a truth value 'False' or F. For example, in mathematics, the statement (2) above is assigned the truth value F. Some statements contain quantifiers, the statement (1) contains the quantifier "All", and the statement (4) contains the quantifier "Some". The statement (5) is a conditional statement of the form:

If statement then statement

The statement (2) contains the word "not", and is the negation of the statement: "20 is less than 30". The quantifier "all", denoted by the symbol "$\forall$", is called the universal quantifier, and is read as "for all". The quantifier "some", denoted by the symbol "$\exists$" is called the existential quantifier; and is read as "there exists". The part of the sentence (1) "are mortal" is a predicate. If we denote this predicate by M, then by $M(x)$ we mean "x is mortal". The statement (1) would then translate in symbolic form as

$$(\forall x)(M(x))$$

which is read as: "for all x, x is mortal", where x is a variable that will vary over the set of all men.

The negation of the statement (1) is the statement: "All men are not mortal". This is equivalent to the statement, of course, to a mathematician, "there exists a man who is not mortal". This, in symbolic form, is written as

$$(\exists x)(\sim M(x))$$

where the symbol "$\sim$" is read as "not". Now read the above symbolic expression as: "there exists a man who is not mortal". Thus, while taking the negation of a statement, the universal quantifier is replaced by the existential quantifier, and the negation of the predicate is taken. Again consider the statement: "There exists an integer which is less than 10". Let us denote the predicate "is less than" in the statement by "LT". Then the expression $LT(x, 10)$ will be read as "x is less than 10". So the statement in symbolic form is given by

$$(\exists x)(LT(x, 10))$$

and read as: "there exists an x such that x is less than 10", where x is a variable that varies over the set of integers. The negation of the above statement is given by

$$(\forall x)(\sim LT(x, 10))$$

which will read as: "for all x, x is not less than 10".

Thus, in general, while taking negation of statement having quantifiers the universal quantifier is replaced by the existential quantifier, the existential quantifier is replaced by the universal quantifier, and the negation of the predicate is taken. If a statement P has the truth value T, then its negation $\sim P$ is given the truth value F and conversely, if P has the truth value F, then $\sim P$ is given the truth value T.

In the statement $(\forall x)(M(x))$, $M(x)$ is called the scope of the quantifier $\forall x$. In general, the scope of a quantifier may be made up of several statements combined by connectives "negation", "and", "or", "conditional", "biconditional" etc. which are denoted by $\sim$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, respectively; there are other connectives as well with which we shall not be concerned here. Thus, the new (compound) statements are generated by repeatedly applying binary connectives $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ between any two statements, and the $\sim$ which is unary connective, that is, applied to a single statement. Thus, if P and Q are statements, then $P \wedge Q$, $P \vee Q$, $P \Rightarrow Q$ (read "If P, then Q"), $P \Leftrightarrow Q$ (read: "P if and only if Q"), $\sim P$ (read "not P"), $\sim (P \Leftrightarrow Q)$, $(P \wedge Q) \vee Q$, etc. are some of the generated statements. A binary connective must always be used between two statement. Thus, the sequences of symbols $P \wedge \wedge Q$, $P \vee \vee Q$, $P \Rightarrow\Rightarrow Q$ are meaningless, but $\sim (\sim P)$ is a valid generated statement. The statement $P \wedge Q$ has the truth value T if both P and Q have the truth values T. $p \vee Q$ has truth value T if either P has the truth value T, or Q has the truth value T. The statement $P \Rightarrow Q$ is given the truth value T if Q has the truth value T wherever P has the truth value T. Thus, the statement $P \Rightarrow Q$ is true if P is a false statement. The statement $P \Leftrightarrow Q$ has the truth value T if both $P \Rightarrow Q$ and $Q \Rightarrow P$ have the truth value T. The statement $(\forall x)(M(x))$ is true if for all values of x, $M(x)$ is true. Similarly, the statement $(\exists x)(LT(x, 10))$ is true if there exists a value of x for which $LT(x, 10)$ is true. The negation of some of the statements are given below:

$$\begin{aligned} \sim (\sim P) &= P, \\ \sim (P \wedge Q) &= (\sim P) \vee (\sim Q), \\ \sim (P \vee Q) &= (\sim P) \wedge (\sim Q), \\ \sim (P \Rightarrow Q) &= \sim ((\sim P) \vee Q) = P \wedge (\sim Q). \end{aligned}$$

By the first equality $\sim (\sim P) = P$ in the above, we mean that the statement $\sim (\sim P)$ in the left of the equality has the same truth value as the statement P in the right of the equality. The two statements are said to be equivalent if they have the same truth values. Thus, $\sim (\sim P)$ and P are equivalent statement.

In the conditional statement $P \Rightarrow Q$, P is called antecedent (hypothesis) statement, and Q is called consequent (conclusion) statement. The conditional

$$\text{if } x^2 < 0 \text{ then } x < -10$$

is a true statement since the antecedent statement: $x^2 < 0$, is given the truth value false. Such a statement is said to be vacuously true.

Example 1.6.1 As a further illustration, we write the negation of continuity at a point of a real valued function of real variable. Recall that a real valued function of real variable is said to be continuous at a point c if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that for each point $x \in \text{dom}(f)$

$$|x - c| < \delta \Rightarrow |f(x) - f(c)| < \epsilon.$$

When all the quantifiers involved are written explicitly, the above statement is equivalent to the following statement:

$$(\forall\epsilon)(\forall x)(\exists\delta)(|x-c|<\delta \Rightarrow |f(x)-f(c)|<\epsilon)$$

where the variable ϵ and δ vary in the set of positive real numbers and the variable x varies in the domain of the function f. The negation of the above statement is

$$(\exists\epsilon)(\exists x)(\forall\delta)(\sim (|x-c|<\delta \Rightarrow |f(x)-f(c)|<\epsilon)).$$

The statement $|x-c|<\delta \Rightarrow |f(x)-f(c)|<\epsilon$ is a conditional statement, and so

$$\sim (|x-c|<\delta \Rightarrow |f(x)-f(c)|<\epsilon) = |x-c|<\delta \wedge (\sim (|f(x)-f(c)|<\epsilon)).$$

Thus, by the negation of the continuity at a point of the function f, we mean

$$(\exists\epsilon)(\exists x)(\forall\delta)(|x-c|<\delta \wedge (\sim (|f(x)-f(c)|<\epsilon)))$$

which will read as: A real valued function of real variables is not continuous at a point c if there exists a real number $\epsilon>0$ such that for each real number $\delta>0$ there exists a point x in the domain of f for which $|x-c|<\delta$ but it is not the case that $|f(x)-f(c)|<\epsilon$.

The contrapositive of a conditional statement "If P then Q" is defined to be the conditional statement "If not Q, then not P", that is, the contrapositive of the statement $P \Rightarrow Q$ is the statement $\sim Q \Rightarrow \sim P$. For example, the contrapositive of the statement: "If the two objects are distinct, then the distance between them is greater than zero" is the conditional statement: "If the distance between two objects is not greater than zero, then the two objects are not distinct". A conditional statement $P \Rightarrow Q$ and its contrapositive both have same truth values, that is, they are equivalent statements for any assignment of truth values to the statements P and Q. As a second example, the contrapositive of the statement: "If $x<0$, then $x^3<0$" is the conditional statement: "If $x^3 \not< 0$, then $x \not< 0$". The equivalence of the $P \Rightarrow Q$ and its contrapositive $\sim Q \Rightarrow \sim P$ is shown below:

P	Q	$P \Rightarrow Q$	$\sim Q$	$\sim P$	$\sim Q \Rightarrow \sim P$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

When P is T, and Q is F which is the second row of the above table both $P \Rightarrow Q$ and $\sim Q \Rightarrow \sim P$ have the truth value F. In rest of the cases, both $P \Rightarrow Q$ and $\sim Q \Rightarrow \sim P$ have the truth values T. Most often, in mathematics, we prove that the conditional statement $\sim Q \Rightarrow \sim P$ is true to prove that the statement $P \Rightarrow Q$ is a true statement.

The converse of the conditional statement $P \Rightarrow Q$ is the conditional statement $Q \Rightarrow P$.

For example, the converse of the statement: "If x is even, then x^2 is even" is the statement: "If x^2 is even, then x is even".

EXERCISE

1. Write the contrapositive and converse of the following statements:
   (a) $(A \subseteq B) \wedge (B \subseteq C) \Rightarrow A \subseteq C$, where A, B and C are sets.
   (b) $(\forall x)(x < 0 \Rightarrow x^2 - x > 0)$, where x denotes a real number.
   (c) If Mohan is ill, then he will not come to the school.
   (d) If a function is differentiable, then it is continuous.
   (e) If a person is a mathematician, then he or she can add two and two.
2. Translate the following in symbolic form and write their negations:
   (a) If a set A is contained in a set B, and B is contained in another set C, then A is contained in C.
   (b) If a tomato is red, then it is ripe.
   (c) All cows in India are of black, white or brown colour.
   (d) Some integers are not even.
   (e) There is a set which contains no element.
   (f) All integers are rational numbers.
   (g) For any integer y, there is an integer x which is less than y.
   (h) Between any two distinct real numbers, there is another real number.
   (i) If an integer is odd, then it is not divisible by two.
   (j) An integer is even if and only if it is the sum of two integers.
3. Let $\mathcal{A}$ be a non-empty collection of sets. Determine the truth value of each of the following statements and their converses.
   (a) $x \in \bigcup_{A \in \mathcal{A}} A \Rightarrow x \in A$ for some $A \in \mathcal{A}$.
   (b) $x \in \bigcup_{A \in \mathcal{A}} A \Rightarrow x \in A$ for every $A \in \mathcal{A}$.
   (c) $x < 0 \Rightarrow x^2 > 0$, where x is a real number.
   (d) $(x \notin A \Rightarrow x \notin B) \Rightarrow B \subseteq A$, where A and B are sets.

1.7 Partially Ordered Sets and Lattices

There is another relation on sets known as partial order relation which is studied extensively in mathematics.

Definition 1.7.1 A partial order relation, denoted by $\leq$, on a set P is a relation satisfying the following properties.

1. $x \leq x$ for all x in P, (Reflexivity)

2. $x \leq y$, and $y \leq x$ implies that $x = y$, (Antisymmetry)

3. $x \leq y$, and $y \leq z$ implies that $x \leq z$. (Transitivity)

The ordered pair $(P, \leq)$ is called a partially ordered set; we sometimes write P when the partial order on P is understood.

Examples 1.7.2

1. The relation "is less than or equal to, denoted by $\leq$" in the set of real number $\mathbb{R}$ is a partial order; infact, the notion of the partial order relation is a generalization of the properties of this relation $\leq$. In this section, the set of real number $\mathbb{R}$ is always equipped with this partial order.

2. The set inclusion relation "is a subset of", denoted by $\subseteq$, on the power set $\mathbb{P}(X)$ of a set X is a partial order.

3. Let P be the set of real valued mappings defined on a set X. The relation "$f \leq g$ if $f(x) \leq g(x)$ for all x in X" is a partial order on P. Notice that the relation $\leq$ in $f(x) \leq g(x)$ above is the usual partial order "is less than or equal to" on $\mathbb{R}$ in (1) above.

4. If $\mathbb{Z}$ is the set of positive integers, then the relation "$x \leq y$ if x divides y" is a partial order.

5. The equality relation "=" is a partial order on all sets.

Let $(P, \leq)$ is a partially ordered set. We write $x < y$ if $x \leq y$ and $x \neq y$. Notice that the relation $<$ is a transitive relation. The relation $<$ is antisymmetric also. For, if $x < y$, and $y < x$, then we would have $x \leq y$ and $x \neq y$, and $y \leq x$ which would give a contradiction in the form $x = y$ and $x \neq y$. We write $y \geq x$ if $x \leq y$. The relation $\geq$ is called the inverse relation of the relation $\leq$. Two elements x, y in P are said to be comparable if either $x \leq y$ or $y \leq x$. There are partial orders on sets, which have the additional property that any two elements are comparable. A partially ordered set in which any two elements are comparable is called a totally ordered set and the partial order is then known as total order relation or linear order. For example, the set $\mathbb{Z}_+$ of positive integers equipped with the usual order relation $\leq$ is a totally ordered set. However, the relation of inclusion "is subset of" on $\mathbb{P}(A)$ is not a total order if the set A contains more than one element. A totally ordered set is sometimes called a chain.

Let $(P, \leq)$ be a partially ordered set, and A be a subset of P. An element u in P is said to be an upperbound of A, (lower bound of A) if $x \leq u$ (respectively, $u \leq x$) for all x in A. An upper bound u of A is called least upper bound (l.u.b.), or supremum (sup A) if $u \leq v$ for all upper bounds v of A. Similarly, a lower bound u of A is called a greatest lower bound (glb) or infimum of A (inf A) if $v \leq u$ for all lower bounds v of A. It is easily seen that least upper bound of set is unique if it exists. Similarly, greatest lower bound of a set A is unique if it exists. For example, in $\mathbb{R}$, define $A =]0, 1[$, then sup $A = 1$ and inf $A = 0$. If $A = \{0, 1\}$ is a subset of $\mathbb{R}$, then sup $A = 1$, and inf $A = 0$; If $A = \emptyset$, then neither sup A nor inf A exist. However, if we consider the partially ordered set to be the subset $P = [0, 1]$ of $\mathbb{R}$, then sup $\emptyset = 0$ and inf $\emptyset = 1$. As an another example, consider the partially ordered set of Example 1.7.2.(2); if $\mathcal{A} = \{A_i\}_{i \in \Lambda}$ is an indexed family of subsets of X, then sup $\mathcal{A}$ is the union $\bigcup_{i \in \Lambda} A_i$ and inf $\mathcal{A}$ is the intersection $\bigcap_{i \in \Lambda} A_i$.

If $(P, \leq)$ is a partially ordered set, and $A = \{a, b\}$ is a two element subset of P, then sup A is denoted by $a \vee b$ (read: "a join b"), and inf A is denoted by $a \wedge b$ (read: "a meet b"). Thus, in Example 1.7.2(2) for any two sets A and B in $\mathbb{P}(X)$, $A \vee B$ is the union $A \cup B$ and $A \wedge B$ is equal to the intersection $A \cap B$. The join $f \vee g$ and meet $f \wedge g$ of f and g in P of Example 1.7.2(3) are given by

$$(f \vee g)(x) = \max\{f(x), g(x)\}, \quad (f \wedge g)(x) = \min\{f(x) = \min\{f(x), g(x)\}. \tag{1.2}$$

An element x in a partially ordered set $(P, \leq)$ is called a maximal element of P if for all y in P, $y \geq x$ implies that $y = x$, that is, no element other than x itself is greater than or equal to x. Similarly, an element x in P is a minimal element of P if for all y in P, $y \leq x$ implies that $y = x$. A partially ordered set may not have any maximal or minimal element, or it may have several of them. The partially ordered set of Example 1.7.2(4) has the unique minimal elememt 1 but no maximal element. The partially ordered set $\mathbb{P}(X)$ of Example 1.7.2 (2) has the unique maximal element X and the unique minimal element the empty set.

If A is a subset of $\mathbb{R}$, and sup A belongs to A, then sup A is often called the maximum of A; Similarly, if inf $A \in A$, then inf A is called the minimum of A.

In $\mathbb{R}$, any non-empty subset which has an upper bound in $\mathbb{R}$ has its supremum in $\mathbb{R}$ (see Theorem 10.3.8). This property of $\mathbb{R}$ is known as order completeness property of $\mathbb{R}$.

Using the above order completeness property of $\mathbb{R}$, it can be shown that a non-empty subset of $\mathbb{R}$ which has a lower bound has its infimum in $\mathbb{R}$.

It may be noted that in the set of rational numbers $\mathbb{Q}$ with usual partial order $\leq$, a non-empty subset having an upper bound may not have its supremum in $\mathbb{Q}$. For example, consider the set $A = \{x \in \mathbb{Q} : x^2 < 2\}$. Then A has neither the supremum nor the infimum in $\mathbb{Q}$.

If $(P, \leq)$ is a partially ordered set, and an element $u \in P$ is such that $x \leq u$ for all x in P, then u is called the greatest element. Similarly, if $v \in P$ is such that $v \leq x$ for all x in P, then v is called the least element. The set $\mathbb{N}_0$ of non-negative integers with usual order $\leq$ has the least element 0 but no greatest element. The same set, but this time with inverse ordering of $\leq$, has the greatest element 0 but no least element.

Definition 1.7.3 A lattice is a partially ordered set $(P, \leq)$ such that $a \vee b \in P$ and $a \wedge b \in P$ for all a, b in P.

Examples 1.7.2(1–5) are all lattices. In Example 1.7.2(4), $a \wedge b$ is the greatest common divisor of a and b, and $a \vee b$ is the least common multiple of a and b.

Definition 1.7.4 Let $(p, \leq)$ be a lattice. A sublattice of P is a subset A of P such that $a \wedge b$ and $a \vee b$ are in A for all a, b in A.

Example 1.7.5 A partially ordered set $(P, \leq)$ in which P is finite can be depicted on the plain paper as follows. If $(a, b) \in \leq$, then b is called cover of a, or a is said to be covered by b if there is no c in P distinct from a and b such that $(a, c) \in \leq$ and $(c, b) \in \leq$; we plot a one level below b, and join a to b by a straight line. We delete from $\leq$ all the reflexive order pairs, that is, pairs of the form (a, a), and the pairs that can be obtained using transitivity, that is, if $(a, b) \in \leq$ and $(b, c) \in \leq$, then (a, c) is deleted from P. Denoting the set of remaining ordered pairs in $\leq$ by $\leq^*$, we plot all the elements in P in such a way that if $(a, b) \in \leq^*$, then a is plotted one level below b, and a and b are joined by a straight line. Thus, if

$P = \{0, a, b, c, d, e, 1\}$ then the relation

$$\leq = \{(0,0),(a,a)(b,b),(c,c),(d,d),(e,e),(1,1),(0,a),(0,b),(a,c),(a,d),(b,d),$$
$$(b,e),(c,1),(d,1),(e,1),(0,c),(0,1),(0,d),(0,e),(a,1),(b,1)\}$$

is a partial order on P. Deleting reflexive and transitive ordered pairs, we obtain $\leq^* = \{(0,a),(0,b),(a,c),(a,d),(b,d),(b,e),(c,1),(d,1),(e,1)\}$. The element 0 is not a cover of any element, so it is plotted at the lowest level. The elements a and b are covers of 0. So a and b are plotted at the second level and so on. The partially ordered set $(P, \leq)$ is depicted in Fig. 1.1.

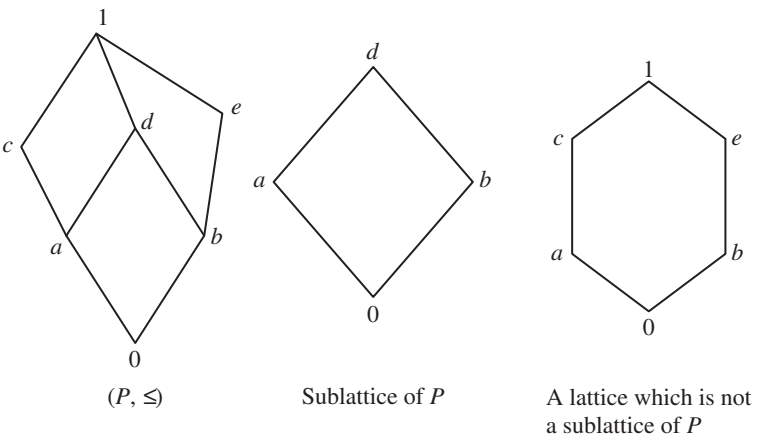


Fig. 1.1

The partially ordered set $(P, \leq)$ is a lattice: $0 \vee a = a$, $0 \wedge a = 0$, $a \vee b = d$, $a \wedge b = 0$, $0 \vee c = c$, $0 \wedge c = 0$, etc. The second diagram in the above figure is a sublattice of P. The third diagram is a lattice in its own right but it is not a sublattice of P. Why?

Example 1.7.6 The set $\mathbb{B}[a, b]$ of all real valued functions defined on the closed interval $[a, b]$ forms a lattice with the partial order: $f \leq g$ if $f(x) \leq g(x)$ for all x in $[a, b]$. The meet $f \wedge g$ and join $f \vee g$ are given by

$$(f \wedge g)(x) = \min\{f(x), g(x)\}, \qquad (f \vee g)(x) = \max\{f(x), g(x)\}$$

for all x in $[a, b]$. The subset $\mathbb{C}[a, b]$ of $\mathbb{B}[a, b]$ consisting of all continuous functions is a sublattice of $\mathbb{B}[a, b]$.

Definition 1.7.7 A lattice P is called complete if for any subset A of P, sup A exists.

It can be shown that if P is a complete lattice, then for any subset A of P, inf A exists.

By induction, it is clear that a finite non-empty subset of a lattice has a supremum and an infimum.

$\mathbb{R}$ is not a complete lattice. The power set $\mathbb{P}(X)$ of an arbitrary set X with inclusion ordering is a complete lattice. The supremum of a collection from $\mathbb{P}(X)$ is its union, and infimum is its intersection.

The study of lattices is a well-developed area of mathematics. For more information on lattices, the reader is referred to Birkoff [5] and Gratzer [21].

Definition 1.7.8 Let A and B be two sets with total order relations $\leq_A$ and $\leq_B$, respectively. A mapping $f : A \to B$ is said to be

(i) monotonically increasing or order preserving if for any $x, y \in A$ $x \leq_A y$ implies that $f(x) \leq_B f(y)$.

(ii) strictly monotonically increasing if for any $x, y \in A$, $x <_A y$ implies that $f(x) <_B f(y)$.

(iii) monotonically decreasing if for any $x, y \in A$, $x \leq_A y$ implies that $f(x) \geq_B f(y)$.

(iv) strictly monotonically decreasing if for any $x, y \in A$, $x <_A y$ implies that $f(x) >_B f(y)$.

We sometimes say that f is a monotone mapping when f is either (strictly) monotonically increasing or a (strictly) monotonically decreasing mapping.

Examples 1.7.9

1. Any constant mapping from $\mathbb{R}$ to $\mathbb{R}$ with the usual order, $\leq$ on $\mathbb{R}$ is a monotone mapping.

2. The identity mapping $i : \mathbb{R} \to \mathbb{R}$, where $\mathbb{R}$ is equipped with the usual order, is a strictly monotonically increasing mapping.

3. If the open interval $]-1, 1[$ and $\mathbb{R}$ both are equipped with the usual order, then the mapping $f :]-1, 1[\to \mathbb{R}$ given by $f(x) = \frac{x}{(1-x^2)}$ is an order preserving bijection.

Definition 1.7.10 Let $(X, \leq_1)$ and $(Y, \leq_2)$ be two partially ordered sets. Then a mapping $f : X \to Y$ such that

$$x \leq_1 y \quad \text{implies that} \quad f(x) \leq_2 f(y)$$

is called a homomorphism of X into Y. If f is a bijective homomorphism, and f^{-1} is also a homomorphism, then f is called an isomorphism between X and Y (or isomorphic mapping from X onto Y), and X and Y are said to be isomorphic.

Definition 1.7.11 A totally ordered set X is said to be well-ordered if every non-empty subset A of X has an infimum which belongs to A. In this case the infimum of A is often called the smallest element of A or the first element of A.

Examples 1.7.12

1. The set of real number $\mathbb{R}$ is totally ordered but $\mathbb{R}$ is not well-ordered.

2. The set of natural numbers $\mathbb{N}$ with usual order $\leq$ is well-ordered.

3. The set of integers $\mathbb{Z}$ with usual ordering $\leq$ is totally ordered, but $\mathbb{Z}$ is not a well-ordered set.

4. The intervals $[0, 1]$, $]0, 1[$, $[0, 1[$, $]0, 1]$ are all totally ordered, but not well-ordered sets.

5. The set $N' = \{1, 2, 3, \ldots, 1^*, 2^*, \ldots\}$ $(n^* > m)$ is a well-ordered set.

Now we introduce the concept of an ordered field.

Definition 1.7.13 Let $(F, +, \cdot)$ be a field, and $\leq$ be a linear order on F such that

(i) $x \leq y$ implies that $x + z \leq y + z$, and

(ii) $x \leq y$, and $0 \leq z$ implies that $xz \leq yz$.

Then $(F, +, \cdot)$ is called an ordered field, denoted by $(F, +, \cdot, \leq)$.

Definition 1.7.14 An ordered field $(F, +, \cdot, \leq)$ is called an order complete field, if every non-empty subset of F which is bounded above has its supremum in F. The real numbers form an order complete field with usual addition, multiplication and order "less than or equal to" (see Chapter 10).

EXERCISE

1. Prove that in the set of natural number $\mathbb{N}$, the relation R defined by aRb if a divides b, that is, $b = ma$ for some $m \in \mathbb{N}$ is a partial order relation.

2. Show that in the set $\mathbb{C}$ of all complex number, the relation $x+iy \leq x_1+iy_2$ if $x_1 < x_2$, or $x_1 = x_2$ and $y_1 \leq y_2$ is a partial order. Is this a total order. Is $\mathbb{C}$ order complete with this order.

3. For two partially ordered sets $(A, \leq_1)$ and $(B, \leq_2)$, the lexicographical ordering (or dictionary ordering) is defined by $(a, b) \leq (c, d)$ if $a <_1 c$, or ($a = c$ and $b \leq_2 d$). If the ordering $\leq_1$ on A and $\leq_2$ on B are total, then the ordering $\leq$ on $A \times B$ is also total. Also if $\leq_1$ and $\leq_2$ are well orderings, then $\leq$ on $A \times B$ is a well-ordering.

4. For partially ordered sets $(A, \leq_1)$ and $(B, \leq_2)$, let $(a, b) \leq (c, d)$ if $a \leq c$ and $b \leq d$. Show that the ordering $\leq$ on $A \times B$ is a partial ordering. If A and B contain more than one element, show that $A \times B$ is never totally ordered by $\leq$.

5. Give an example of a partially ordered set which is not a lattice.

1.8 Number System

Natural Numbers

The description of natural numbers, integers, rational numbers, real numbers and complex numbers though is not exactly necessary for the purpose of the present book, and most books in the area do not opt for this section, we include the brief description of the numbers primarily to refresh the memory of the reader as many examples later in the text will be constructed out of these numbers. Every one of us is aware of natural numbers $0, 1, 2, 3, \ldots$ which are used for counting, but still there is a need that these be defined in a proper manner. What really are natural numbers? There are at least two ways to define them. The following axiomatic definition of natural numbers is due to Peano (1858–1932).

Peano Axioms for Natural Numbers. There exists a set $\mathbb{N}_0$ of element called natural numbers satisfying the following axioms:

1. There exists an element in $\mathbb{N}_0$ which we denote by 0.

2. For each element $n \in \mathbb{N}_0$, there exists a unique element n^+ in $\mathbb{N}_0$ called the successor of n.
3. For all $n \in \mathbb{N}_0$, $n^+ \neq 0$.
4. If $n, m \in \mathbb{N}_0$ and $n^+ = m^+$, then $n = m$.
5. If P is a subset of $\mathbb{N}_0$ such that $0 \in P$, and $n \in P$ implies $n^+ \in P$, then $P = \mathbb{N}_0$.

We write $0^+ = 1$, $1^+ = 2$, $2^+ = 3$, $3^+ = 4$, and so on.

The axiom (5) of Peano's axioms is the basis for the principle of mathematical induction which is stated as follows. If $P(n)$ is a mathematical statement for each $n \in \mathbb{N}_0$, and is such that

(i) $P(0)$ is true,

(ii) $P(n^+)$ is true whenever $P(n)$ is true.

Then $P(n)$ is true for each natural number n in $\mathbb{N}_0$.

The relationship between this principle and axiom (5) can be understood in the following manner. Let A be the set of natural numbers n for which the statement $P(n)$ is true. The number 0 is in A since $P(0)$ is true, and the second condition of the mathematical induction implies that n^+ is in A whenever n is in A. By axiom (5), $A = \mathbb{N}_0$, that is $P(n)$ is true for every natural number n.

The addition of natural numbers is defined inductively as follows:

(i) $0 + y = y$,

(ii) $x^+ + y = (x + y)^+$.

Using (1) and (2) above and the induction, the following basic properties of natural numbers can be easily proved.

1. $x + (y + z) = (x + y) + z$. (Associativity)
2. $x + y = y + x$. (Commutativity)
3. $x + z = y + z$ implies that $x = y$. (Cancellation law)

The multiplication of natural numbers is defined inductively:

(i) $0y = 0$,

(ii) $x^+y = xy + y$.

The following properties are consequences of the above definition and of the induction.

1. $x(yz) = (xy)z$. (Associativity)
2. $xy = yx$. (Commutativity)
3. If $z \neq 0$, then $xz = yz$ implies that $x = y$. (Cancellation law)
4. $x(y + z) = xy + xz$. (Distributivity)

The concept of order relation in $\mathbb{N}_0$ can be defined as follows: A natural number x is said to be greater than a natural number y, denoted by $x > y$, if there exists a natural number $z \neq 0$ such that $x = y + z$.

From the above definition, the following properties of the relation $>$ can be proved.

1. $x > y$, and $y > z$ implies that $x > z$. (Transitive law)

2. For any x and y in $\mathbb{N}_0$, one and only one of the following holds

$$x > y, \qquad x = y, \qquad y > x. \qquad \text{(Trichotomy law)}$$

We close this section with the theorem about the set of natural numbers which asserts that the set of natural numbers $\mathbb{N}$ is unique upto one-to-one correspondence.

Theorem 1.8.1 Let $\mathbb{N}$ and $\mathbb{N}'$ be two sets satisfying Peano's axioms (1)–(5). Then there exists a one-to-one correspondence $f : \mathbb{N} \to \mathbb{N}'$ such that $f(0) = 0'$ and $f(n^+) = f(n)^+$, where 0 and $0'$ are the elements of Peano's axiom (1) for the set $\mathbb{N}$ and $\mathbb{N}'$, respectively. Such a mapping is called an isomorphism.

In view of the above theorem, the set $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ both can be considered as the sets of natural numbers, and both the sets are popular in mathematics.

Integers

The way to construct the set of integers is to regard it as the set of all differences between the elements of ordered pairs of natural numbers. For instance

$(2, 3)$ gives rise to -1
$(2, 4)$ gives rise to -2
$(1, 1)$ gives rise to 0.
$(3, 2)$ gives rise to 1.
$(50, 20)$ gives rise to 30, etc.

Formally: Let $a, b, c, d \in \mathbb{N}_0$. We say that the ordered pair (a, b) is related to (c, d), written $(a, b) \sim (c, d)$, if $a + d = b + c$.

It is easily verified that the relation $\sim$ is an equivalence relation, and therefore partitions the set $\mathbb{N}_0 \times \mathbb{N}_0$ into equivalence classes. An equivalence class is called an integer. For example, the equivalence class of the ordered pair $(0, 1)$, denoted by $[0, 1]$, is given by

$$[0, 1] = \{(a, b) \in \mathbb{N}_0 \times \mathbb{N}_0 : (a, b) \sim (0, 1), \quad \text{that is,} \quad a + 1 = b + 0\}.$$

The integer $[0, 1]$ is denoted by -1 (read: "minus 1") in our traditional notation of the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

In general, the equivalence class of $(a, b) \in \mathbb{N}_0 \times \mathbb{N}_0$ is denoted by $[a, b]$. Intuitively, we intend that the equivalence class $[a, b]$ of the ordered pair (a, b) represents the integer $a - b$. With this understanding, we define the addition and the multiplication operation on the set of integers as follows: Let $a, b, c, d \in \mathbb{N}_0$, then

(i) $[a, b] + [c, d] = [a + c, b + d]$, (Addition)

(ii) $[a, b] \times [c, d] = [ac + bd, ad + bc]$. (Multiplication)

The operations of the addition and the multiplication as defined above are operations on equivalence classes, and as such we need to check that these are well defined. We show that the addition is well defined, and leave the multiplication to the reader. For addition to be well defined, it is needed to be shown that the result of two equivalence classes does not depend on the pairs of natural numbers which are chosen to represent these equivalence classes. That means we must verify that if $[p, q] = [a, b]$, and $[r, s] = [c, d]$, then

$$[p, q] + [r, s] = [a, b] + [c, d].$$

Since $(p, q) \sim (a, b)$, and $(r, s) \sim (c, d)$,

$$p + b = q + a, \quad \text{and} \quad r + d = s + c.$$

Then $$\begin{aligned}(a + c) + (q + s) &= (a + q) + (c + s)\\ &= b + p + d + r\\ &= (b + d) + (p + r).\end{aligned}$$

Therefore, $[a + c, b + d] = [p + r, q + s]$ as required. It can be seen that the addition and the multiplication satisfy commutative, associative and distributive laws. Also notice that for $(a, b) \in \mathbb{N}_0 \times \mathbb{N}_0$,

$$\begin{aligned}[a, b] + [0, 0] &= [a, b],\\ [a, b] \times [1, 0] &= [a, b],\\ [a, b] \times [0, 0] &= [0, 0].\end{aligned}$$

Thus, the integer $[0, 0]$ is the additive identity, and is denoted by 0, and the integer $[1, 0]$ behaves like a multiplicative identity, and it is denoted by 1. Also for any $a, b \in \mathbb{N}_0$,

$$[a, b] + [b, a] = [0, 0].$$

In view of the above property, we denote $[b, a]$ by $-[a, b]$ and abbreviate $[a, b] + (-[c, d])$ by $[a, b] - [c, d]$. Thus, the subtraction operation '–' is introduced as a legitimate operation on integers.

We denote the set of integers by $\mathbb{Z}$, and use variables $x, y, \ldots$ to denote the integers.

Now we introduce an order on $\mathbb{Z}$.

An integer $[a, b]$ in $\mathbb{Z}$ is said to be positive if $b < a$ (as elements of $\mathbb{N}_0$). Again it must be shown that the notion of positive integer is well defined. For this, we need to show that if $p, q \in \mathbb{N}_0$, and $[p, q] = [a, b]$, then $q < p$. Since $(p, q) \sim (a, b)$, $p + b = q + a$. Since $b < a$, it follows using the property of $<$ in $\mathbb{N}_0$ that $q < p$ as required. The set of positive integers is denoted by $\mathbb{Z}_+$. The order "is less than" denoted by $<$ on $\mathbb{Z}$ is defined as

$$x < y \quad \text{if} \quad y - x \in \mathbb{Z}_+.$$

By the order relation $x \leq y$ (read: "x is less than or equal to y") we mean that $x < y$ or $x = y$.

We can define the modulus (or absolute value) function as

$$|x| = \begin{cases} x & \text{if } x \geq 0,\\ -x & \text{if } x < 0.\end{cases}$$

Let x and y be integers. Then there exist integer q and r such that

$$x = qy + r, \quad \text{where} \quad 0 \leq r < |y|.$$

This fact is known as Euclidean algorithm.

Consider the set $S = \{[n, 0] : n \in \mathbb{N}_0\}$. The element $[0, 0]$ in S behaves like a zero in Peano's axiom (1). The mapping $f : S \to S$ given by $f([n, 0]) = [n^+, 0]$ acts like a successor mapping in Peano's axioms. Then it is easily seen that S satisfies all the axioms of Peano. Therefore, $\mathbb{N}_0$ and S are isomorphic, the isomorphism associates each $n \in \mathbb{N}_0$ with $[n, 0] \in S$. So S has the same mathematical structure as $\mathbb{N}_0$; we identify $\mathbb{N}_0$ with S. It is in this sense that natural numbers are regarded as integers.

Rational Numbers

A similar process as used to define integers from natural numbers is used to define rational numbers from integers. It should be noted that out of the four arithmetic operations, viz. addition, multiplication, subtraction and division, for natural numbers addition and multiplication are defined, and the integers have three operations : addition, multiplication and subtraction. As an ordered pair of natural numbers represents basically the difference of the natural numbers in the ordered pair, a rational number is introduced as an ordered pair of integers which essentially represents the quotient of integers in the ordered pair. Thus, the ordered pair $(-1, 2)$ essentially defines a rational number usually denoted by $-\frac{1}{2}$. As there are several ordered pairs that represent the same rational number, we define an equivalence relation on the set of all ordered pairs of integer, that is, on the set $\mathbb{Z} \times \mathbb{Z}$, that clubs together all such ordered pairs that represent the same rational number. Now we begin in a formal manner.

Let $a, c \in \mathbb{Z}$ and $b, d \in \mathbb{Z} - \{0\}$. The ordered pair (a, b) is said to be related to (c, d), denoted $(a, b) \sim (c, d)$, if $ad = bc$.

It is easily seen that the above relation $\sim$ is an equivalence relation on $\mathbb{Z} \times \mathbb{Z} - \{0\}$ and so it partitions the set $\mathbb{Z} \times \mathbb{Z} - \{0\}$ into equivalence classes. A rational number is an equivalence class in this partition. For integers $a \in \mathbb{Z}$, $b \in \mathbb{Z} - \{0\}$, the equivalence class of (a, b) under the relation $\sim$ is denoted by a/b. For example, the equivalence class of the pair $(1, 2)$ is denoted by $1/2$. The set of all rational numbers is denoted by $\mathbb{Q}$.

Now we define the arithmetic operations on the set of rational numbers. Let $a, c \in \mathbb{Z}$ and $b, d \in \mathbb{Z} - \{0\}$. Then the addition is defined by

$$a/b + c/d = (ad + bc)/bd$$

and the multiplication by

$$a/b \times c/d = ac/bd.$$

Notice that ad is the multiplication of integers a and d, and the addition $ad + bc$ is the addition of integers ad and bc.

It should be remarked that the above definitions of the addition and the multiplication of rational numbers are in the same spirit as our earlier intuitively learned basis of these operations for rational numbers:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b} + \frac{c}{d} = \frac{ac}{bd}.$$

This is what we tell a beginner.

As in the case of integers, here also there is a need to prove that the above operations are well-defined. Again the well-definedness of the addition operation is proved, and the multiplication is left to the reader. Also, always in a rational number $a/b, b \neq 0$. Now let $a/b = p/q$ and $c/d = r/s$. We are to show that $a/b + c/d = p/q + r/s$. Since $a/b = p/q$, $(a, b) \sim (p, q)$ and similarly, $(c, d) \sim (r, s)$. Therefore,

$$aq = bp, \quad \text{and} \quad cs = dr.$$

Multiplying the first equality by ds and the second by bq, and adding, we get

$$aqds + csbq = bpds + drbq.$$

This gives the equality $(ad + bc)qs = bd(ps + qr)$ as required. Next we list some of the properties of rational numbers, and introduce the operations of subtraction and division.

1. If $a/b \in \mathbb{Q}$, and $x \in \mathbb{Z} - \{0\}$, then $ax/bx = a/b$.

2. The addition and the multiplication of rational numbers are commutative, associative and the multiplication distributes over the addition.

3. The rational number $0/1$ is the additive identity (zero) of rational numbers, that is,

$$a/b + 0/1 = a/b.$$

   Also
$$a/b \times 0/1 = 0/1.$$

4. The rational number $1/1$ is the multiplicative identity of rational numbers, that is,

$$a/b \times 1/1 = a/b.$$

5. For any $a, b \in \mathbb{Z} - \{0\}$, we have

$$a/b \times b/a = ab/ab = 1/1.$$

   The division by a/b of a rational number p/q is defined to be the multiplication of p/q by the multiplicative inverse b/a of a/b, that is

$$(p/q)/(a/b) = b/a \times p/q.$$

   Notice, of course, that we can do this if b/a is a rational number, that is, only if $a \neq 0$.

6. For $a, b \in \mathbb{Z}$, and $b \neq 0$, we have

$$a/b + (-a)/b = (ab - ab)/b^2 = 0/b^2 = 0/1.$$

   Also note that $(-a)/b = a/(-b)$. We write $(-a)/b$ as $-a/b$. Thus, $-a/b$ is an additive inverse of a/b, and this enables us to define the subtraction operation. Let $a, b, c, d \in \mathbb{Z}$ with $b \neq 0, d \neq 0$. We define

$$a/b - c/d = a/b + (-c)/d = a/b + (-c/d)\,.$$

Now we define an order relation on $\mathbb{Q}$.

We say that an element a/b of $\mathbb{Q}$ is positive if $ab > 0$. This notion of positive rational numbers is well defined. For, let $a/b = c/d$, and a/b is positive. Then $(a, b) \sim (c, d)$. This means that $ad = bc$. Thus, the following equalities follow.

$$cdb^2 = bcbd = adbd = abd^2.$$

Since $b^2 > 0$, $d^2 > 0$, and $ab > 0$, we have, $cd > 0$, as required. The set of positive rational numbers is denoted by $\mathbb{Q}_+$. Now we define the order relation "is less than", denoted by $<$, on $\mathbb{Q}$ by

$$x < y \quad \text{if} \quad y - x \in \mathbb{Q}_+.$$

Also, $x \leq y$ if $x < y$ or $x = y$.

As the set of natural numbers $\mathbb{N}_0$ can be regarded as a subset of the set of integers, in the same way $\mathbb{Z}$ can be regarded as a subset of $\mathbb{Q}$. Infact, the rational numbers of the form $a/1$ behave as integers. For instance,

$$a/1 + b/1 = (a + b)/1,$$

$$a/1 \times b/1 = (ab)/1.$$

If $S = \{a/1 : a \in \mathbb{Z}\}$, then there is an isomorphism between $\mathbb{Z}$ and S that maps a to $a/1$. Thus, we can regard $\mathbb{Z}$ as a subset of $\mathbb{Q}$, and forget the distinction between $a \in \mathbb{Z}$ and $a/1 \in \mathbb{Q}$.

There are certain properties the order $<$ on $\mathbb{Q}$ has in relation to various arithmetic operations.

Theorem 1.8.2 1. Given $x \in \mathbb{Q}$, one of the following holds

$$\text{(a) } x \in \mathbb{Q}_+, \quad \text{(b) } x = 0, \quad \text{(c) } -x \in \mathbb{Q}_+.$$

2. If $x, y \in \mathbb{Q}_+$, then $x + y \in \mathbb{Q}_+$, and $x \times y \in \mathbb{Q}_+$.

Proof. 1. Let $x = a/b$. If $a = 0$, then $x = 0/b = 0$, where 0 on the right denotes the rational number zero. If $a, b \in \mathbb{Z}_+$, then $ab \in \mathbb{Z}_+$, and so $a/b \in \mathbb{Q}_+$. If $a \in \mathbb{Z}_+$ and $-b \in \mathbb{Z}_+$, then $a(-b) \in \mathbb{Z}_+$, and so $(-ab) \in \mathbb{Z}_+$. That is, $-x \in \mathbb{Q}_+$. In the other cases we argue similarly.

2. Let $x = a/b$, and $y = c/d$, where $b \neq 0$, $d \neq 0$, $ab > 0$, and $cd > 0$. Then

$$x + y = a/b + c/d = (ad + bc)/bd.$$

Now $(ad + bc)bd = abd^2 + b^2cd > 0$. Therefore, $x + y \in \mathbb{Q}_+$. The case of product is left to the reader.

Notice that the set of rational numbers with the operations of addition and multiplication is a field. Also, it can be verified that our set $\mathbb{Z}$ of integers with the operations of addition and multiplication defined on it is an integral domain. The property that a field has, which an integral domain does not necessarily have, is that every non-zero element has a multiplicative inverse. The field of rational numbers is so developed that it meets the above requirement. Indeed, a field can be developed starting from an arbitrary integral domain with some trivial modifications such that the given integral domain can be embedded in this field. We do not pursue this line any further. This will, however, be an interesting exercise for the reader.

It is also clear that $(\mathbb{Q}, +, \times, [0,0], [1,0], \leq)$ is an ordered field.

The order relation $<$ on $\mathbb{Q}$ has the following properties which are of fundamental importance in mathematical analysis.

Theorem 1.8.3 The order relation $<$ on $\mathbb{Q}$ is dense, that is, if $x, y \in \mathbb{Q}$ and $x < y$, then there exists a rational number z such that $x < z$ and $z < y$.

Proof. Let $x = a/b$, and $y = c/d$ be such that $x < y$. Then $(bc - ad)/bd \in \mathbb{Q}_+$. Now

$$x + y = (ad + bc)/bd.$$

Let us take $z = (ad + bc)/2bd$. Then

$$\begin{aligned} z - x &= (ad + bc)/2bd - a/b = (adb + b^2c - 2abd)/2b^2d \\ &= b/2b \times (bc - ad)/bd \\ &= 1/2 \times (bc - ad)/bd \end{aligned}$$

which is in $\mathbb{Q}_+$ since $1/2 \in \mathbb{Q}_+$. Thus, $x < z$. Similarly, we can prove that $z < y$.

Theorem 1.8.4 The order $<$ on $\mathbb{Q}$ is Archimedean, that is, if $x, y \in \mathbb{Q}_+$, then there exists an $n \in \mathbb{Z}$ such that $y < nx$.

Proof. Let $x = a/b$, and $y = c/d$ be elements of $\mathbb{Q}_+$. Then $ab > 0$, and $cd > 0$. So a and b have the same sign, and also c and d have the same sign. Then one of the following two cases arise:

1. $ad \in \mathbb{Z}_+$, and $bc \in \mathbb{Z}_+$.
2. $-ad \in \mathbb{Z}_+$, and $-bc \in \mathbb{Z}_+$.

In case (1) by Euclidean algorithm, there exist q and r in $\mathbb{Z}_+$ such that

$$bc = q(ad) + r \quad \text{and} \quad 0 \leq r < ad.$$

Consequently, $bc < q(ad) + ad$, that is, $bc < (q + 1)ad$. Now taking n to be $q + 1$, we have $nad - bc \in \mathbb{Z}_+$ as required. Similarly, we can argue the case (2). The proof is now complete.

In the remainder of the text, we can afford to write $\frac{a}{b}$ for a rational number a/b.

Real Numbers

The mathematical limitations of rational numbers were known at least since the time of Pythagoras as it is he who is credited with the first proof that there is no rational number whose square is 2. Interpreted in geometrical terms, this means that the length of the diagonal of a square of a unit side is not a rational number. The problem then is how to describe the length of this diagonal. This is a particular case of a more general limitation of $\mathbb{Q}$. A polynomial equations with integer coefficients may not have a solution in $\mathbb{Q}$. In the above example, the equation involved is $x^2 - 2 = 0$. In general, a quadratic equation $ax^2 + bx + c = 0$, with $a, b, c \in \mathbb{Z}$ does not have a solution in $\mathbb{Q}$ unless $b^2 - 4ac$ is a perfect square. The rational numbers suffer another limitation: A cauchy sequence of rational

numbers need not converge in $\mathbb{Q}$. A popular example of this sort is the sequence defined by

$$x_1 = 1,$$
$$x_{n+1} = \frac{4 + 3x_n}{3 + 2x_n} \quad \text{for all } n \in \mathbb{Z}_+.$$

A third limitation of the rational numbers is that a non-empty and bounded from above subset of rational numbers need not have its supremum in $\mathbb{Q}$. Hence, there is a need to have a system of numbers in which we can overcome the above limitations of the rational numbers. Mathematicians have struggled for a long time in developing such a system of numbers which is known as the real number system.

There are mainly three processes for the introduction of real numbers from the rational numbers. These are chains of segments due to Cantor, Dedekind cuts of Dedekind, and Cauchy sequences due to Cauchy. We describe here the first two, that is, the definition of a real number via the process of chains of segments and Dedekind-cuts. First, we explain Cantor's approach with the help of an example. Consider the inequalities:

$$1 < \sqrt{2} < 2$$
$$1.4 < \sqrt{2} < 1.5$$
$$1.41 < \sqrt{2} < 1.42$$
$$1.414 < \sqrt{2} < 1.415$$
$$1.4142 < \sqrt{2} < 1.4143$$
$$\vdots \quad \vdots \quad \vdots$$

In each row, the number $\sqrt{2}$ lies between a pair of rational numbers. For example, in the third row, $\sqrt{2}$ lies between 1.41 and 1.42; the pair $(1.41, 1.42)$ is called a segment. Thus, we get a sequence of segments: $\langle (1,2), (1.4, 1.5), (1.41, 1.42), (1.414, 1.415), \ldots \rangle$; This sequence, called a chain of segments defines $\sqrt{2}$. Thus, $\sqrt{2}$ is determined by an infinite process. The above chain of segments $\langle (1,2), (1.4, 1.5), (1.41, 1.42), \ldots \rangle$ satisfies certain properties which we formulate explicitly in the following.

A sequence $\langle (a_n, b_n) \rangle$ of segments of rational numbers is called a chain if

(i) $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$ for all $n \in \mathbb{N}$, and

(ii) $b_n - a_n \leq \frac{1}{2^n}$ for all $n \in \mathbb{N}$.

The condition (2) in the preceding definition gives the speed of convergence of the chain. This is not necessary, but some other condition for the convergence of the chain has to be given, and this one happens to be rather simple.

We now define an equivalence relation in the set of these chains. Two chains $\langle (a_n, b_n) \rangle$ and $\langle (c_n, d_n) \rangle$ of segments of rational numbers are said to be related, denoted by $\langle (a_n, b_n) \rangle \sim \langle (c_n, d_n) \rangle$, if $c_n \leq b_n$ and $a_n \leq d_n$ for all n in $\mathbb{N}$. Intuitively, $\langle (a_n, b_n) \rangle \sim \langle (c_n, d_n) \rangle$ if the segment (a_n, b_n) meets (c_n, d_n); there are two situations:

$$\begin{array}{cccc} * & \times & * & \times \\ a_n & c_n & b_n & d_n \end{array} \qquad \begin{array}{cccc} \times & * & \times & * \\ c_n & a_n & d_n & b_n \end{array}$$

It can be easily seen that the relation $\sim$ in the above definition is an equivalence relation, and thus partitions the set of all chains of segments of rational numbers into equivalence classes. A real number is an equivalence class. To return to our example, we define $\sqrt{2}$ as the real number representing the equivalence class of the chain $\langle(1,2),(1.4,1.5),(1.41,1.42),\ldots\rangle$.

Next we define a real number as a Dedekind-cut, the Cauchy construction of a real numbers is postponed to Chapter 10. A Dedekind-cut is a non-empty proper subset x of $\mathbb{Q}$ such that

(a) x is "closed downward", that is

$$p \in x \quad \text{and} \quad q < p \quad \text{implies that } q \in x,$$

(b) x has no largest element.

A Dedekind-cut is called a real number, and the set of real numbers is denoted by $\mathbb{R}$.

The ordering in $\mathbb{R}$ is particularly simple. For x and y in $\mathbb{R}$, x "is less than" y, denoted by $x < y$, if x is a proper subset of y, that is,

$$<= \{(x,y) \in \mathbb{R} \times \mathbb{R} : x \subset y\},$$

We say that x "is less than or equal to" y, denoted by $x \leq y$, if $x < y$ or $x = y$. Similarly, x is greater than y, denoted by $x > y$, if $y < x$, and x is greater than or equal to y, denoted by $x \geq y$, if $x > y$ or $x = y$.

For the ordering $<$ on $\mathbb{R}$, the following properties are true.

1. $x < y$ and $y < z$ implies that $x < z$. (Transitivity)
2. The law of trichotomy holds, that is, exactly one of the following alternatives is true

$$x < y, \quad x = y, \quad y < x.$$

The addition operation for $\mathbb{R}$ is defined in terms of the addition operation in $\mathbb{Q}$

$$x + y = \{p + q : p \in x, \text{ and } q \in y\}.$$

Notice that $x + y \in \mathbb{R}$. Also

$$x + (y + z) = (x + y) + z, \quad \text{(Associativity)}$$

$$x + y = y + x. \quad \text{(Commutativity)}$$

The zero element 0 of $\mathbb{R}$ is defined to be the set of negative rational numbers

$$0 = \{p \in \mathbb{Q} : p < 0\}.$$

It is easily seen that 0 is a real number, and that

$$x + 0 = x.$$

A real number x is said to be positive if $x > 0$. The set of positive real numbers is denoted by $\mathbb{R}_+$.

The negative of a real number x, denoted by $-x$, is defined as

$$-x = \{r \in \mathbb{Q} : \text{there exists } s \text{ in } \mathbb{Q} \text{ such that } s > r, \text{ and } -s \notin x\}.$$

Thus, the difference '$-$' is introduced as a legitimate operation in $\mathbb{R}$

$$x - y = x + (-y).$$

The cancellation law holds, that is

$$x + z = y + z \text{ implies that } x = y.$$

It may be observed that the addition operation preserves the order:

$$x < y \text{ if and only if } x + z < y + z.$$

Theorem 1.8.5 For any $x \in \mathbb{R}$, one of the following holds:

$$x \in \mathbb{R}_+, \quad x = 0, \quad -x \in \mathbb{R}_+.$$

Theorem 1.8.6 The ordering $<$ of $\mathbb{R}$ is dense, that is, for any $x, y \in \mathbb{R}$ such that $x < y$, there exists a z in $\mathbb{R}$ such that $x < z$, and $z < y$.

Theorem 1.8.7 The ordering $<$ on $\mathbb{R}$ is Archimedean, that is, given x, y in $\mathbb{R}_+$, there exists a positive integer r such that $y < rx$.

Notice that the ordering $\leq$ is a partial ordering on $\mathbb{R}$. The following theorem, which is known as "the ordering completeness property of $\mathbb{R}$ or least upper bound property of $\mathbb{R}$" holds.

Theorem 1.8.8 Any non-empty subset of $\mathbb{R}$, which is bounded above has a least upper bound in $\mathbb{R}$.

The above theorem has several applications in mathematical analysis. For example, it is used to show that a continuous real valued function defined on a closed interval assumes its maximum, and this, in turn, is used to prove the mean value theorems of calculus.

The absolute value or modulus of a real number x, denoted by $|x|$, is defined as

$$|x| = \max\{x, -x\}.$$

We may be tempted to define the product of the real numbers x and y as

$$\{pq : p \in x \text{ and } q \in y\}$$

in analogy to the definition of $x + y$. If we take this as the definition of product, then the product of two negative real numbers x and y is not a real number. Therefore, we use the following variation of the above idea.

If x and y are two non-negative real numbers, that is, $x \geq 0$, $y \geq 0$, then the product of x and y, denoted by $x.y$, is defined as

$$x.y = 0 \cup \{pq : 0 \leq p \in x \text{ and } 0 \leq s \in y\}$$

if x and y are both negative real numbers, then the product of x and y is defined as

$$x.y = |x|.|y|$$

and if one of the real numbers x and y is negative and the other is non-negative, then the product is given by

$$x.y = -(|x|.|y|).$$

The product $x.y$ is written simply as xy when no confusion arises.

The real number $\{p \in \mathbb{Q} : p < 1\}$ is denoted by 1. Clearly, $0 < 1$. The following properties of multiplication and addition operations can now be verified.

1. xy is a real number.
2. (i) $x(yz) = (xy)z$. (Associative law)
   (ii) $xy = yx$. (Commutative law)
   (iii) $x(y + z) = xy + xz$. (Distributive law)
   $(x + y)z = xz + yz$.
3. If $0 < z$, then
   $x < y$ if and only if $xz < yz$.

The multiplicative inverse of a real number $x > 0$, denoted by x^{-1} or $\frac{1}{x}$, is defined as

$$x^{-1} = \frac{1}{x} = \{r \in \mathbb{Q} : r \leq 0 \text{ or, } r > 0 \text{ and } \frac{1}{r} \notin x \text{ but } \frac{1}{r} \text{ is not the supremum of } x\}$$

and if $x < 0$, then $x^{-1} = -(|x|)^{-1}$.

Once the multiplication is introduced, it is straightforward to introduce the operation of division on $\mathbb{R}$. Let x and y be real numbers and $y \neq 0$. The division of x by y, denoted by $\frac{x}{y}$, is defined by

$$\frac{x}{y} = xy^{-1}.$$

Thus, we have now completed the process of introducing the four fundamental operations of arithmetic, viz, addition, multiplication, subtraction and division.

The foregoing results indicate that $(\mathbb{R}, +, \cdot, 0, 1, \leq)$ forms an ordered field. Since Theorem 1.8.8 holds, $(\mathbb{R}, +, \cdot, 0, 1, \leq)$ is an ordered complete field. Notice that the analogous statement of Theorem 1.8.8 in $\mathbb{Q}$: "Any non-empty bounded from above subset of $\mathbb{Q}$ has a least upper bound in $\mathbb{Q}$" is not true. For example, the set $\{r \in \mathbb{Q} : r^2 < 2\}$ has no least upper bound in $\mathbb{Q}$. Due to this, the ordered field $(\mathbb{Q}, +, \cdot, 0, 1, \leq)$ is not an ordered complete field.

It can be shown that any two ordered complete fields are isomorphic to each other, that is, there is, a bijective correspondence f between them that preserves addition, multiplication, and the order (see Chapter 10), that is:

$$f(x + y) = f(x) + f(y),$$
$$f(xy) = f(x)f(y),$$
$$x \leq y \text{ implies that } f(x) \leq f(y).$$

Therefore, upto isomorphism the ordered complete field $(\mathbb{R}, +, \cdot, 0, 1, \leq)$ is unique.

There is an embedding function E from $\mathbb{Q}$ into $\mathbb{R}$ that assigns to each rational number r in $\mathbb{Q}$ the real number

$$E(r) = \{q \in \mathbb{Q} : q < r\}.$$

For this embedding function E, the following theorem holds.

Theorem 1.8.9 The embedding function $E : \mathbb{Q} \to \mathbb{R}$ satisfies the following properties.

1. E is injective.
2. E preserves the addition: $E(r + s) = E(r) + E(s)$.
3. E preserves the multiplication: $E(rs) = E(r)E(s)$.
4. $E(0) = 0$ and $E(1) = 1$.
5. $r < s$ if and only if $E(r) < E(s)$.

Here the first occurrence of 0 in (4) is the additive identity in $\mathbb{Q}$, and the second occurrence of 0 is the additive identity in $\mathbb{R}$, and similarly for the occurences of 1.

In view of the above theorem, we identify the rational number r with the real number $E(r)$ and forget any distinction between r and $E(r)$. It is in this sense the field of rational numbers is considered as a subfield of $\mathbb{R}$.

Decimal Notation

We conclude the discussion of real numbers by describing the decimal representation of them. In our daily mathematical experience, we use the decimal expression for real numbers, for example

$$1.41,\ 5.52324\ldots,\ 2.424242\ldots,\ 5.576582\ldots$$

This notation has certain limitations, for instance, in the second, third and fourth examples above, we were not able to specify completely which real numbers are represented by them. However, in the third decimal expression, we observe some pattern, viz, the digits 4 and 2 are being repeated; it is called non-terminating repeating decimal. The second and third are examples of non-terminating non-repeating decimals; usually, it is more difficult to do algebra with such types of numbers; for example, consider the addition or multiplication of non-terminating non-repeating decimals. The first decimal expression 1.41, of course, completely represents a real number which is usually known as a terminating decimal. Such a number can be considered as non-terminating repeating decimal by appending a sequence of zeros: $1.4100000\ldots$.

Let $n \in \mathbb{N}_0$, and let $a_1, a_2, \ldots$ be digits in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Consider the expression

$$n.a_1a_2\ldots a_n\ldots$$

The above expression represents the sum of the series

$$n + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots$$

Equivalently, the expression can be said to represent the "limit" of the sequence

$$n,\ n + \frac{a_1}{10},\ n + \frac{a_1}{10} + \frac{a_2}{10^2},\ n + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3}, \cdots$$

The sequence defined above is a Cauchy sequence of rational numbers, and the real number determined by this sequence is represented by the decimal expression $n.a_1a_2a_3\ldots$ (see Chapter 10).

Thus, it follows that to each decimal expression there corresponds a real number. It requires a little more effort to show the converse, that is, that each real number in $\mathbb{R}$ may be represented by a decimal expression.

Theorem 1.8.10 Every positive real number may be represented uniquely by an expression $n.a_1a_2\ldots$, where $n \in \mathbb{N}_0$, and each a_i is an integer with $0 \leq a_i \leq 9$, and where there is no $m \in \mathbb{N}_0$ such that $a_i = 9$ for all $i > m$ (that is, the sequence $a_1, a_2, \ldots$ is not to end with the infinite sequence of 9's).

Proof. Let x be a positive real number. By the Archimedean property, there is an $r \in \mathbb{Z}_+$ such that $r.1 > x$. Take r to be least such number, and let $n = r - 1$ so that n is the greatest integer less than or equal to x. Therefore, we can write $x = n + a$ for some real number a such that $0 \leq a < 1$. Then $0 \leq 10a < 10$. Let a_1 be the largest integer less than or equal to $10a$. Then $10a = a_1 + r_1$, where $r_1 \in \mathbb{R}$, and $0 \leq r_1 < 1$. Then $a = \frac{a_1}{10} + \frac{r_1}{10}$. Repeating the above process with number $10r_1$, we get $10r_1 = a_2 + r_2$, and consequently $x = n + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{r_2}{10^3}$, where $r_2 \in \mathbb{R}$ and $0 \leq r_2 < 1$. This process goes on, possibly indefinitely, and generates the digits $a_1, a_2, \ldots$ of the required decimal expression.

The above process cannot lead to a repeating 9. Suppose, if possible, there exists an $n_0 \in \mathbb{N}_0$ such that $a_n = 9$ for all $n > n_0$. If $n_0 = 0$, then $a = .999\ldots$, so that

$$10a = 9.999\ldots$$

This gives $a = 1$ a contradiction. Thus, $n_0 > 0$. Then $r_{n_0-1} = .99\ldots$. Then

$$10r_{n_0} = 9.999\ldots = 9 + r_{n_0}.$$

This again gives $r_{n_0} = 1$, a contradiction.

Finally, we demonstrate uniqueness. Suppose that $n.a_1a_2\ldots = n'.b_1b_2\ldots$ be two decimal expression for a real number x. Now

$$\sum_{i=1}^{\infty} \frac{a_i}{10^i} < \sum_{i=1}^{\infty} \frac{9}{10^i} = 1$$

since the digit 9 is not repeating. Similarly, $\sum_{i=1}^{\infty} \frac{b_i}{10^i} < 1$. Hence, the largest integer less than or equal to x is n and also n', so that $n = n'$. Let k be the smallest integer such that $a_k \neq b_k$, and suppose without loss of generality that $a_k < b_k$. Then

$$\begin{aligned}
0 &= \sum_{i=1}^{\infty} \frac{a_i}{10^i} - \sum_{i=1}^{\infty} \frac{b_i}{10^i} \\
&= \sum_{i=k}^{\infty} \frac{a_i}{10^i} - \sum_{i=k}^{\infty} \frac{b_i}{10^i} \\
&= \frac{(a_k - b_k)}{10^k} + \sum_{i=k+1}^{\infty} \frac{a_i}{10^i} - \sum_{i=k+1}^{\infty} \frac{b_i}{10^i} \\
&< -\frac{1}{10^k} + \frac{1}{10^k} = 0
\end{aligned}$$

since $a_k - b_k \leq -1$, and

$$\sum_{i=k+1}^{\infty} \frac{a_i}{10^i} < \sum_{i=k+1}^{\infty} \frac{9}{10^i} = \frac{1}{10^k}$$

where we have used the fact that the sequence $a_1, a_2, \ldots$ does not have a tail of repeating 9, that is, there is no $n \in \mathbb{N}_0$ such that $a_i = 9$ for all $i > n$. Thus, we get the contradiction $0 < 0$. So, $a_k = b_k$ for all k, and the proof is complete.

Remark. 1. It may be remarked that the expression $0.99\ldots$ represents the sum of the series

$$\frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \cdots$$

which is equal to

$$\frac{9}{10}.\left(1 + \frac{1}{10} + \frac{1}{10^2} + \cdots\right) = 1.$$

Consequently $1.00\ldots$ and $0.99\ldots$ are representations of the same real number. Similarly, $1.500\ldots$ and $1.4999\ldots$ represent the same real number.

2. If we apply the construction of the above theorem to the number $-\pi$, we would obtain $n = -4$ (the greatest integer less than or equal to $-\pi$), and $a_1 = 8$, $a_2 = 5$, $a_3 = 8$ etc. It is perfectly reasonable way to obtain the decimal expression of a negative real number. However, the normal representation of $-\pi$ is $-3.14159\ldots$.

Theorem 1.8.11 A real number has a decimal expression which terminates or recurs if and only if it is a rational number.

Proof. Since a terminating decimal expression can be written as a non-terminating expression with recurring 0, we consider only recurring case. Let x be a rational number. Then, in the proof of Theorem 1.8.10, a is also a rational number. Let $a = \frac{p}{q}$ with $p, q \in \mathbb{N}_0$, and $q \neq 0$. Then

$$a = \frac{p}{q} = \frac{a_1}{10} + \frac{r_1}{10}.$$

Therefore,

$$0 \leq r_1 = \frac{(10p - a_1 q)}{q} < 1.$$

Denoting $10p - a_1 q$ by p_1, we have

$$r_1 = \frac{p_1}{q}, \text{ where } 0 \leq p_1 < q.$$

Similarly,

$$r_2 = \frac{p_2}{q}, \text{ where } 0 \leq p_2 < q$$

and so on. There must exists an i such that $p_i = p_j$ for some $j < i$, that is, $r_i = r_j$. From this point on, the process will repeat, and the sequence $a_j, a_{j+1}, \ldots, a_{i-1}$ will recur in the decimal expression for x.

Conversely, suppose that x has a recurring decimal expression $x = n.a_1a_2 \ldots a_r a_{r+1} \ldots a_{r+s} \ldots$, where the sequence $a_{r+1}, a_{r+2}, \ldots, a_{r+s}$, recurs. Then

$$\begin{aligned} x &= n + \frac{a_1}{10} + \cdots + \frac{a_r}{10^r} + \frac{a_{r+1}}{10^{r+1}} + \cdots + \frac{a_{r+s}}{10^{r+s}} + \frac{a_{r+s+1}}{10^{r+s+1}} + \cdots + \frac{a_{r+2s}}{10^{r+2s}} + \cdots \\ &= n + \frac{a_1}{10} + \cdots + \frac{a_r}{10^r} + \left(\frac{a_{r+1}}{10^{r+1}} + \cdots + \frac{a_{r+s}}{10^{r+s}}\right) \\ &\qquad + \frac{1}{10^s}\left(\frac{a_{r+1}}{10^{r+1}} + \cdots + \frac{a_{r+s}}{10^{r+s}}\right) + \frac{1}{10^{2s}}(\,) + \cdots \\ &= n + \frac{a_1}{10} + \cdots + \frac{a_r}{10^r} + \left(\frac{a_{r+1}}{10^{r+1}} + \cdots + \frac{a_{r+s}}{10^{r+s}}\right)\left(1 + \frac{1}{10^s} + \frac{1}{10^{2s}} + \cdots\right) \\ &= n + \frac{a_1}{10} + \cdots + \frac{a_r}{10^r} + \left(\frac{a_{r+1}}{10^{r+1}} + \cdots + \frac{a_{r+s}}{10^{r+s}}\right) / \left(1 - \frac{1}{10^s}\right) \end{aligned}$$

which is a rational number.

Remarks. The number 10 occurs in decimal expression as historical fact. We can use the integers 2, 3, 4, ... as the basis of the real numbers. Much of the development will remain the same except some trivial modifications. In particular, with base 2, the expression for a real number of the form $n.a_1a_2 \ldots$, where $n \in \mathbb{N}_0$, and $a_i \in \{0, 1\}$ is called a binary expression.

Complex Numbers

There is no real number x which satisfies the equation $x^2 + 1 = 0$, that is, the equation has no real roots. In order to discuss the possible roots of the above equation, the need was felt to extend the real member system $\mathbb{R}$ in some way. Swiss mathematician Euler (1707–1783) was the first to introduce the symbol i for $\sqrt{-1}$ so that $i^2 = -1$. In this way, i is the root of the equation $x^2 + 1 = 0$. He called the symbol i imaginary. Subsequently, the symbol $a + ib$, where a and b are real numbers, was introduced, and it was called a complex number. However, we now introduce a complex number as a pair of real numbers, and express it in the form $a + ib$.

An ordered pair (a, b) of real numbers a, b is called a complex number. Thus, $(2, 3)$, $(.1, \sqrt{5})$, $(\sqrt{3}, 0)$ are all complex numbers. The set of complex numbers is denoted by $\mathbb{C}$.

The two complex numbers (a, b) and (c, d) are said to be equal, denoted $(a, b) = (c, d)$, if $a = c$, and $b = d$.

The addition and multiplication on $\mathbb{C}$ are defined as follows.

$$\begin{aligned} (a, b) + (c, d) &= (a + c, b + d). \\ (a, b)(c, d) &= (ac - bd, bc + ad). \end{aligned}$$

With these operations, $\mathbb{C}$ satisfies all the axioms of a field. $(0, 0)$ and $(1, 0)$ are identities for the addition and multiplication, respectively. The additive inverse of (a, b) is $(-a, -b)$, and the multiplicative inverse of a non-zero element (a, b) is (c, d), where

$$c = \frac{a}{(a^2 + b^2)} \quad \text{and} \quad d = \frac{-b}{(a^2 + b^2)}.$$

The multiplicative inverse of (a, b) is denoted by $(a, b)^{-1}$.

The difference of two complex numbers (a, b) and (c, d) is defined as the addition of (a, b) with the additive inverse of (c, d), that is

$$(a, b) - (c, d) = (a, b) + (-c, -d).$$

The division of (a, b) by a non-zero complex number (c, d) is defined to be the multiplication of (a, b) by the multiplicative inverse of (c, d), that is

$$\begin{aligned}\frac{(a,b)}{(c,d)} = (a,b)(c,d)^{-1} &= (a,b)\left(\frac{c}{(c^2+d^2)}, \frac{-d}{(c^2+d^2)}\right)\\ &= \left(\frac{ac}{(c^2+d^2)} + \frac{bd}{(c^2+d^2)}, \frac{bc}{(c^2+d^2)} - \frac{ad}{(c^2+d^2)}\right)\\ &= \left(\frac{(ac+bd)}{(c^2+d^2)}, \frac{(bc-ad)}{(c^2+d^2)}\right).\end{aligned}$$

Thus, the four fundamental arithmetic operations on $\mathbb{C}$ are introduced. However, $\mathbb{C}$ does not have an order structure similar to that of $\mathbb{R}$.

The mapping $f : \mathbb{R} \to \mathbb{C}$ defined by $f(a) = (a, 0)$ is a field isomorphism of $\mathbb{R}$ into $\mathbb{C}$ so that $\mathbb{R}$ can be considered as a subfield of $\mathbb{C}$ if we identify the real number a with $(a, 0)$. If we put $i = (0, 1)$, then

$$(a, b) = (a, 0) + (0, b) = (a, 0) + (0, 1)(b, 0) = a + ib.$$

Thus, if we adopt the convention of writing a for $(a, 0)$ and i for $(0, 1)$ and vice-versa, then the complex number (a, b) can be written as $a + ib$. It is in this sense that the notation $a + ib$ is used in place of the complex number (a, b).

We write $z = a + ib$, a is called real part of z and b the imaginary part of z, written as $a = \mathrm{Re}\, z$, and $b = \mathrm{Im}\, z$.

If $z = a + ib$, then the absolute value of z, or modulus of z, is defined as

$$|z| = (a^2 + b^2)^{\frac{1}{2}}.$$

The conjugate of $z = a + ib$, $\bar{z}$ is defined to be the complex number $a - ib$, that is $\bar{z} = a - ib$. The following properties are easy to prove.

1. $|z|^2 = z\bar{z}$.
2. If $z \neq 0$, then $\dfrac{1}{z} = \dfrac{\bar{z}}{|z|^2}$.
3. $\mathrm{Re}\, z = \dfrac{(z + \bar{z})}{2}$, $\mathrm{Im}\, z = \dfrac{z - \bar{z}}{2i}$.
4. $\overline{(z_1 + z_2)} = \bar{z}_1 + \bar{z}_2$, $\overline{(z_1 z_2)} = \bar{z}_1 \bar{z}_2$.
5. $|z_1 z_2| = |z_1||z_2|$.
6. $\left|\dfrac{z_1}{z_2}\right| = \dfrac{|z_1|}{|z_2|}$, $z_2 \neq 0$.

7. $|\bar{z}| = |z|$.
8. $|z_1 + z_2| \leq |z_1| + |z_2|$.
9. $||z_1| - |z_2|| \leq |z_1 - z_2|$.
10. $|z_1 + z_2|^2 = |z_1|^2 + 2\,\mathrm{Re}\, z_1\bar{z}_2 + |z_2|^2$.
11. $|z_1 - z_2|^2 = |z_1|^2 - 2\,\mathrm{Re}\, z_1\bar{z}_2 + |z_2|^2$.
12. $|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2\left(|z_1|^2 + |z_2|^2\right)$.

A complex number $z = a + ib$ is associated the point (a, b) in the plane $\mathbb{R}^2$ Thus, there is one-to-one correspondence between the points of $\mathbb{C}$ and the points of $\mathbb{R}^2$. The addition of complex number corresponds to the addition law of vectors in the vector space $\mathbb{R}^2$. Let z and w be in $\mathbb{C}$. Draw the vectors from $0 = (0, 0)$ to z and w. These vectors can be considered as the two sides of the parallelogram with 0, z and w as the three vertices. The fourth vertex of this parallelogram turns out to be $z + w$.

The number $|z - w|$ is the length of the line segment joining z and w, and the following inequality holds:

$$|z_1 - z_2| \leq |z_1 - z_3| + |z_3 - z_2|$$

which, interpreted in geometrical terms, means that the length of a side of the triangle with vertices z_1, z_2, z_3 is less than or equal to the sum of the lengths of the other two sides, and is known as triangle inequality. By induction, we get

$$|z_1 + z_2 + \cdots + z_n| \leq |z_1| + |z_2| + \cdots + |z_n|.$$

Also note that $|z|$ is the length of the line segment joining $0 = (0, 0)$ and z, and the $\bar{z}$ is the reflection of the point z in x-axis (that is, the real axis).

The point $z = a + ib$ has the polar co-ordinates (r, θ), where $a = r\cos\theta$, $b = r\sin\theta$, then $r = |z|$, and θ is the angle between the positive real axis and the line segment from 0 to z. θ is considered positive if measured anti-clockwise and negative if measured clockwise. θ is known as argument of z, written as $\theta = \arg z$. Since θ plus any integral multiple of 2π can be substituted in the equations $a = r\cos\theta$, $b = r\sin\theta$ to get the same values of a and b, arg is not a function. With a slight abuse of notation we denote by $\arg z$ the set of all values of $\arg z$. For each value of $z \neq 0$ there is exactly one value θ in the set $\arg z$ such that $-\pi < \theta \leq \pi$, this θ is then called the principal value of $\arg z$. We write

$$\mathrm{cis}\,\theta = \cos\theta + i\sin\theta.$$

Now if $z_1 = r_1\,\mathrm{cis}\,\theta_1$ and $z_2 = r_2\,\mathrm{cis}\,\theta_2$, then

$$z_1 z_2 = r_1 r_2\,\mathrm{cis}(\theta_1 + \theta_2).$$

Therefore, $\arg(z_1 z_2) = \arg z_1 + \arg z_2$ where $\arg z_1 + \arg z_2 = \{\theta_1 + \theta_2 : \theta_1 \in \arg z_1$ and $\theta_2 \in \arg z_2\}$. Similarly, $\arg(z_1/z_2) = \arg z_1 - \arg z_2$. In particular, if $z = r\,\mathrm{cis}\,\theta$, and n is an integer, then

$$z^n = r^n\,\mathrm{cis}(n\theta)$$

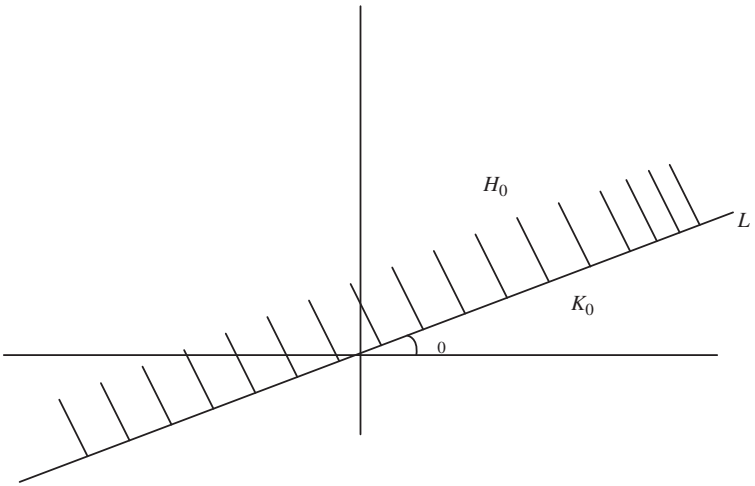


Fig. 1.2

which is known as De Moivre's formula. Given any non-zero complex number a, having argument α the equation $z^n = a$ has n distinct roots given by

$$|a|^{\frac{1}{n}} \operatorname{cis} \frac{(\alpha + 2k\pi)}{n}, \quad k = 0, 1, 2, \ldots, n-1.$$

Thus, for each non-zero complex number a, there are n distinct nth roots of a as given above.

When the co-ordinate plane $\mathbb{R}^2$ is thought of as consisting of complex numbers and is given the algebra of complex numbers including the modulus and complex conjugate as defined above, it is the known as complex plane.

As mentioned earlier, a complex number z can be considered as a vector joining 0 to z. Let $a, b \in \mathbb{C}$, and $b \neq 0$. Then the straight line L in $\mathbb{C}$ passing through a and in the direction of the vector b is defined as

$$L = \{z = a + tb : t \in \mathbb{R}\}.$$

Equivalently,

$$L = \{z : \operatorname{Im} \frac{(z-a)}{b} = 0\}.$$

The line L divides the complex plane in two parts. The sets

$$H_a = \{z : \operatorname{Im} \frac{(z-a)}{b} > 0\}, \quad \text{and} \quad K_a = \{z : \operatorname{Im} \frac{(z-a)}{b} < 0\}$$

are upper and lower half planes respectively. For $a = 0$, the upper and lower half planes are shown in Fig. 1.2.

To show this, without loss of generality, we can take $|b| = 1$. Then $H_0 = \{z : \operatorname{Im}(z/b) > 0\}$.

Assume that $b = \operatorname{cis}\theta_0$, $z = r \operatorname{cis}\theta$. Then $\frac{z}{b} = r \operatorname{cis}(\theta - \theta_0)$. Thus $z \in H_0$ if and only if $\sin(\theta - \theta_0) > 0$ if and only if $\theta_0 < \theta < \pi + \theta_0$. So H_0 is the half plane lying to the left of L if we are "walking along the line L in the direction of b". Now observe that $H_a = H_0 + a = \{z + a : z \in H_0\}$. Thus, H_a is the half plane to the left of L.

EXERCISE

1. Prove that for every natural number n, either $n = 0$, or $n = m^+$ for some natural number m. Hence, show that product of two non-zero natural numbers is non-zero.

2. Prove that for any natural numbers m and n, either $m \leq n$, or $n \leq m$.

3. Let $m, n \in \mathbb{N}_0$ with $m \neq 0$. Prove that there exists $r \in \mathbb{N}_0$ such that $n < rm$.

4. Prove that every non-empty set of integers which is bounded below has a least element.

5. Let $x, y \in \mathbb{Z}$ be such that $xy = 0$. Prove that $x = 0$, or $y = 0$.

6. Let $x \in \mathbb{Q}$. Prove that there is an integer which is greater than x. Prove also that there is an integer which is smaller than x.

7. Prove that for any positive real number x, there exists a positive rational number q with $q < x$.

8. Let $x, y \in \mathbb{R}$ with $x < y$. Prove that there exists $z \in \mathbb{Q}$ such that $x < z$, and $z < y$.

9. Let $x, y \in \mathbb{Q}$ with $x < y$. Prove that there exists $z \in \mathbb{R} - \mathbb{Q}$ such that $x < z$, and $z < y$.

10. Prove that every positive real number has a unique positive square root in $\mathbb{R}$.

11. Use Theorem 1.8.10 to find the decimal expression for the rational numbers $\frac{1}{7}$, $\frac{2}{9}$, $\frac{3}{23}$. Can you suggest any general rule governing the length of the recurring sequence in such expansions.

12. An integer a is said to be divisible by an integer $b \neq 0$ if there exists an integer c such that $a = bc$. The integer b is called a divisor (or factor) of a, and a is called a multiple of b. Show that the relation of divisibility in the set of integers is reflexive, transitive, but not symmetric.

13. An integer $p > 1$ is said to be a prime number if its only divisor are ± 1. Show that the set of prime numbers is infinite.

14. Show that there is no rational number whose square is 12.

15. Prove that no order can be defined in the complex field $\mathbb{C}$ that turns it into an ordered field.

16. Determine the regions of the complex plane specified by

$$\left| \frac{(z - \alpha)}{(1 - \bar{\alpha} z)} \right| < 1, = 1, \text{ or } > 1$$

where α is a complex number such that $|\alpha| < 1$.

1.9 Finite Sets and Infinite Sets

The reader must have encountered finite sets and infinite sets earlier. In this section and the next, we shall learn about them in a precise manner. It will be seen that intuitively looking simple facts about finite sets may require significant amount of work to establish them.

If n is a positive integer, then the set

$$\{x : x \in \mathbb{Z}_+ \text{ and } 1 \leq x \leq n\}$$

denoted $\{1, 2, \ldots, n\}$ is called a section of the positive integers.

Definition 1.9.1 A set A is said to be finite if it is either empty, or if there is a bijective correspondence between the set A and the set $\{1, 2, \ldots, n\}$ for some positive integer n.

When A is empty, A is said to have 0 number of elements, in the case when there is bijection between A and the section $\{1, 2, \ldots, n\}$, A is said to have n elements.

A set is said to be infinite if it is not a finite set.

The number of elements in a finite set A is uniquely determined by A. This fact though appear quite obvious requires a proof, and it involves a number of steps.

Lemma 1.9.2 Let n be a positive integer. Let A be a set. Let a_0 be an element in A. Then $f : A \to \{1, 2, \ldots, n+1\}$ is a bijective correspondence if and only if there exists a bijective correspondence $g : A - \{a_0\} \to \{1, 2, \ldots, n\}$.

Proof. If $g : A - \{a_0\} \to \{1, 2, \ldots, n\}$ is bijective, then define $f : A \to \{1, 2, \ldots, n+1\}$ by mapping $f(x) = g(x)$ for all x in $A - \{a_0\}$ and $f(a_0) = n + 1$. It is now obvious that f is a bijective mapping. Conversely, assume that $f : A \to \{1, 2, \ldots, n+1\}$ is a bijective mapping. If $f(a_0) = n+1$, then the restriction $f_{A-\{a_0\}}$ of f is a bijection between $A - \{a_0\}$ and $\{1, 2, \ldots, n\}$. So we assume that $f(a_0) = m < n + 1$. Then there exists a point a_1 distinct from a_0 such that $f(a_1) = n + 1$. Define a mapping $h : A \to \{1, 2, \ldots, n+1\}$ by

$$\begin{aligned} h(a_0) &= n + 1, \\ h(a_1) &= m, \\ h(x) &= f(x) \quad \text{if} \quad x \in A - \{a_0, a_1\}. \end{aligned}$$

Then it is easily seen that h is a bijection. The restriction $h_{A-\{a_0\}}$ is the desired bijection.

Theorem 1.9.3 Let A be a set, and $f : A \to \{1, 2, \ldots, n\}$ be a bijection for some $n \in \mathbb{Z}_+$. If B is a proper subset of A, then (1) there is no bijection $g : B \to \{1, 2, \ldots, n\}$, (2) there exists a bijection $h : B \to \{1, 2, \ldots, m\}$ for some positive integer $m < n$ provided the set B is non-empty.

Proof. Let Z_0 be a subset of $\mathbb{Z}_+$ consisting of those integers n for which the theorem holds. By induction, we show that $Z_0 = \mathbb{Z}_+$. For $n = 1$, since $f : A \to \{1\}$ is a bijection, A is a singleton of the form $\{a\}$. Then B can only be the empty set. Since there is no bijection between the empty set B and the non-empty set $\{1\}$, $1 \in Z_0$. Now we assume that $n \in Z_0$, and we show that $n + 1$ is in Z_0. So let $f : A \to \{1, 2, \ldots, n+1\}$ be a bijection, and B be a proper subset of A. If B is empty, there is no bijection between B and the set $\{1, 2, \ldots, n+1\}$. So we assume that B is non-empty. Let a_0 be a point in B, and a_1 be a point in $A - B$. By Lemma 1.9.2, there is a bijection

$$g : A - \{a_0\} \to \{1, 2, \ldots, n\}.$$

Since $B - \{a_0\}$ is a proper subset of $A - \{a_0\}$, and the theorem is true for n, it follows that

(i) there is no bijection $h : B - \{a_0\} \to \{1, 2, \ldots, n\}$, and

(ii) there exists a bijection $k : B - \{a_)\} \to \{1, 2, \ldots, m\}$ for some $m < n$ provided $B - \{a_0\} \neq \emptyset$.

Lemma 1.9.2 together with (1) implies that there is no bijection between B and $\{1, 2, \ldots, n+1\}$. To prove the part (2), observe that if $B - \{a_0\} = \emptyset$, then $B = \{a_0\}$, and there is bijection between B and the set $\{1\}$. If $B - \{a_0\} \neq \emptyset$, then by (2) and Lemma 1.9.2, there is a bijection between B and $\{1, 2, \ldots, m+1\}$, where $m + 1 < n + 1$. Therefore, the theorem holds for $n + 1$, and the proof is complete.

Corollary 1.9.4 There is no bijection of a finite set A with a proper subset of A.

Proof. Let B be a proper subset of A, and $f : A \to B$ be a bijection. Now there is a bijection $g : A \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{Z}_+$. Then the composition $g \circ f^{-1} : B \to \{1, 2, \ldots, n\}$ is a bijection. This contradicts Theorem 1.9.3.

Corollary 1.9.5 The number of elements in a finite set A is uniquely determined by A.

Proof. Assume that $m < n$, and $f : A \to \{1, 2, \ldots, n\}$ and $g : A \to \{1, 2, \ldots, m\}$ are two bijections. Then $g \circ f^{-1} : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ is a bijection, contradicting Corollary 1.9.4. The number of elements in the empty set is of course 0 by the definition.

Corollary 1.9.6 If B is a subset of a finite set A, then B is finite. If B is a proper subset of a finite set, then the number of element in B is less than the number of elements in A.

Corollary 1.9.7 $\mathbb{Z}_+$ is not finite.

Proof. The mapping $f : \mathbb{Z}_+ \to \mathbb{Z}_+ - \{1, 2, \ldots, k\}, k \in \mathbb{Z}_+$ defined by $f(n) = n + k$ is a bijection, and therefore the proof follows in view of corollary 1.9.4.

Theorem 1.9.8 Let A be a non-empty set, and $n \in \mathbb{Z}_+$. The following statements are equivalent.

1. $f : \{1, 2, \ldots, n\} \to A$ is a surjective mapping.
2. $g : A \to \{1, 2, \ldots, n\}$ is an injective mapping.
3. A is finite, and has at most n elements.

Proof. (1) implies (2): We define $g : A \to \{1, 2, \ldots, n\}$ by $g(a) =$ smallest element of $f^{-1}(\{a\})$. The set $f^{-1}(\{a\})$ is non-empty for all $a \in A$ since f is a surjective mapping, and $f^{-1}(\{a\})$ has a unique smallest element since the set $\{1, 2, \ldots, n\}$ is a well ordered set (see section 1.11). Also if a and b are any two distinct elements of A, then the sets $f^{-1}(\{a\})$ and $f^{-1}(\{b\})$ are disjoint, and thus have distinct smallest elements. Therefore, g is injective.

(2) implies (3): Let $B = g(A)$. Then the mapping $g' : A \to B$ given by $g'(x) = g(x)$ is bijective. Since $B \subseteq \{1, 2, \ldots, n\}$, there is a bijection $h : B \to \{1, 2, \ldots, p\}$, where $p \leq n$. Then the composition $h \circ g' : A \to \{1, 2, \ldots, p\}$ is a bijection. Therefore, A is finite and has $p\,(\leq n)$ elements.

(3) implies (1): Suppose that A has p, $p \leq n$, elements. Then there is a bijection $h : \{1, 2, \ldots, p\} \to A$. If $p = n$, then h is a required surjection. So assume that $p < n$. In this case, we extend the mapping h by putting $h(m) = h(1)$ for all m such that $p < m \leq n$.

Corollary 1.9.9 Let $\{A_i\}_{i\in J}$ be a finite family of finite-sets. Then the sets $\bigcup_{i\in J} A_i$ and $\prod_{i\in J} A_i$ are finite.

A set A is equinumerous to a set B, written $A \approx B$, if there exists a bijective mapping $f : A \to B$.

Corollary 1.9.10 Any set equinumerous to a proper subset of itself is infinite.

Proof. The proof follows from Corollary 1.9.4.

Example 1.9.11 The open interval $]0, 1[$ is equinumerous to the set $\mathbb{R}$ of real numbers. A geometrical construction of a one-to-one correspondence is shown in Fig. 1.3. In this the open interval $]0, 1[$ is bent into a semicircle with centre C. Each point of $]0, 1[$ is paired with its projection (from C) on the real line. There is also an analytical proof that $]0, 1[\approx \mathbb{R}$. The mapping $f :]0, 1[\to \mathbb{R}$ defined by $f(x) = \tan\left(\frac{\pi(2x-1)}{2}\right)$ is a one-to-one correspondence. Thus, it is quite possible that a set is equinumerous to a proper subset of itself. In view of Corollary 1.9.4, this is never possible for finite sets. Galileo observed in 1638 that $\mathbb{N}_0$ was equinumerous to the set $\{0, 1, 4, 9, \ldots\}$ of squares of natural numbers, and found this to be quite strange. Similarly, the set of even integers is equinumerous to the set of all integers.

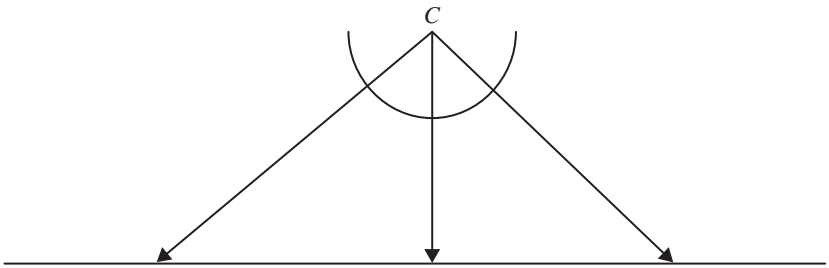


Fig. 1.3

The following theorem holds.

Theorem 1.9.12 For any sets A, B and C, the following holds.

1. $A \approx A$.
2. If $A \approx B$, then $B \approx A$.
3. If $A \approx B$, and $B \approx C$, then $A \approx C$.

However, it is not the case that any two infinite sets are equinumerous.

Theorem 1.9.13 The set $\mathbb{N}_0$ is not equinumerous to the set $\mathbb{R}$ of real numbers.

Proof. If $\mathbb{N}_0 \approx \mathbb{R}$, then there exists a bijective mapping $f : \mathbb{N}_0 \to \mathbb{R}$. For each k, we write the decimal expression of $f(k)$ as

$$x_k.a_{k1}a_{k2}a_{k3}\ldots,$$

where x_k is an integer, and a_{ki} are decimal digits 0, 1, ..., 9. We construct a real number z as

$$z = 0.b_{11}b_{22}b_{33}\ldots\ .$$

such that $b_{kk} = 7$ if $a_{kk} \neq 7$ and $b_{kk} = 6$ if $a_{kk} = 7$. Then z is not in the range of f, a contradiction.

1.10 Countable and Uncountable Sets

The reader has already seen the principle of mathematical induction. This principle is based on the set of natural numbers $\mathbb{N}_0$. What type of set $\mathbb{N}_0$ is? We cannot get a bijective correspondence between $\mathbb{N}_0$ and a section of positive integers. So $\mathbb{N}_0$ is not a finite set, and hence is an infinite set. The following definition introduces a further distinction among sets.

Definition 1.10.1 An infinite set A is said to be denumerable or countably infinite if there is a bijective correspondence between $\mathbb{Z}_+$ and A. A set is said to be countable if it is either finite or denumerable.

Examples 1.10.2

1. The set $\mathbb{Z}_+$ is countably infinite. The identity mapping $i : \mathbb{Z}_+ \to \mathbb{Z}_+$ given by $i(n) = n$ for all n in $\mathbb{Z}_+$ is bijective. Thus, $\mathbb{Z}_+$ is countable.
2. The set $\mathbb{Z}$ of integers is countable. The mapping $f : \mathbb{Z} \to \mathbb{Z}_+$ given by

$$f(n) = \begin{cases} 2n+1 & \text{if } n \geq 0, \\ -2n & \text{if } n < 0 \end{cases}$$

   is a bijection.
3. The set $\{\frac{1}{n} : n \in \mathbb{Z}_+\}$ is a countable set of rational numbers.
4. The set of even natural numbers and the set of odd natural numbers are countable sets.

The following theorem charcaterizes countable sets.

Theorem 1.10.3 Let A be a non-empty set. Then the following statements are equivalent.

1. A is countable.
2. There is a surjective mapping $f : \mathbb{Z}_+ \to A$.
3. There is an injective mapping $f : A \to \mathbb{Z}_+$.

Proof. (1) implies (2): Suppose that A is a countable set. If A is countably infinite, then, by the definition, there is a bijection and hence a surjection $f : \mathbb{Z}_+ \to A$. If A is finite, then there exists bijection $h : \{1, 2, \ldots, n\} \to A$ for some $n \in \mathbb{Z}_+$. This surjection is extended to obtain a surjection $f : \mathbb{Z}_+ \to A$ by defining

$$f(i) = \begin{cases} h(i) & \text{if } 1 \leq i \leq n, \\ h(n) & \text{if } i > n. \end{cases}$$

(2) implies (3): Let $f : \mathbb{Z}_+ \to A$ be a surjection. Then the mapping $g : A \to \mathbb{Z}_+$ defined by

$$g(x) = \inf f^{-1}(x)$$

is the required injection. Note that g is well-defined since $f^{-1}(x)$ is non-empty for all x in A. Also, for $x \neq y$, $f^{-1}(x)$ and $f^{-1}(y)$ are disjoint, and therefore, $\inf f^{-1}(x) \neq \inf f^{-1}(y)$.

(3) implies (1): Let $g : A \to \mathbb{Z}_+$ be an injection. We define $B = g(A)$. Then the mapping $h : A \to B$ defined by $h(x) = g(x)$ is a bijection. To show that A is countable, it is enough to show that B is countable. For this, we show that any subset of $\mathbb{Z}_+$ is countable. Let C be a subset of $\mathbb{Z}_+$. If C is finite, then it is countable by the definition. So let C be an infinite subset of $\mathbb{Z}_+$. We construct a bijection $k : \mathbb{Z}_+ \to C$ using the principle of recursive definition (see Section 1.13) as follows:

$$k(1) = \inf C,$$
$$k(n) = \inf(C - k(\{1, 2, \ldots, n-1\}) \quad \text{for all } n \geq 2.$$

Notice that $C - k(\{1, 2, \ldots, n-1\})$ is non-empty, for if it were empty, then the mapping $k : \{1, 2, \ldots, n-1\}) \to C$ would be a surjection, and hence C would be a finite set. Also if $m < n$, then $k(m) \in k(\{1, 2, \ldots, n-1\})$ which means that $\inf(C - k(\{1, 2, \ldots, m-1\})) < \inf(C - k(\{1, 2, \ldots, n-1\}))$. Therefore, k is injective. Also it shows that $k(\mathbb{Z}_+)$ is infinite. Next we show that k is surjective. Let y be any element in C. Since $k(\mathbb{Z}_+)$ is infinite, there exists an element $n \in \mathbb{Z}_+$ such that $k(n) > y$. Let m be the smallest positive integer such that $k(m) \geq y$. Then $k(i) < y$ for all $i < m$. This means that y does not belong to the set $k(\{1, 2, \ldots, m-1\})$. Therefore, $k(m) = \inf(C - k(\{1, 2, \ldots, m-1\})) \leq y$. Hence, $k(m) = y$.

Example 1.10.4 The set $\mathbb{Q}_+$ of positive rational numbers is denumerable. We arrange the elements of $\mathbb{Q}_+$ in the form of rows and columns as shown below.

$$\begin{array}{ccccccc}
\frac{1}{1} & & & & & & \\
\downarrow & & \frac{1}{2} & \to & \frac{1}{3} & \frac{1}{4} & \cdots \\
\frac{2}{1} & \nearrow & & \swarrow & & & \\
 & & \frac{2}{2} & & \frac{2}{3} & \frac{2}{4} & \cdots \\
\frac{3}{1} & \swarrow & & & & & \\
\downarrow & & \frac{3}{2} & & \frac{3}{3} & \frac{3}{4} & \cdots \\
\frac{4}{1} & & & & & &
\end{array}$$

In the above arrangement, the fractions in the first row have numerator 1, and the denominators in increasing order as $1, 2, 3, \ldots$. Similarly, the second row has 2 in the numerator of each fraction, and the denominators in increasing order as $1, 2, 3, \ldots$, and so on. Notice that each element of $\mathbb{Q}_+$ appears in some row of this arrangement. Some elements of $\mathbb{Q}_+$ are, of course, repeated in the arrangement, for example, $\frac{1}{1}, \frac{2}{2}, \frac{3}{3}, \ldots$ represent the same number $\frac{1}{1}$. If we consider all the element in the arrangement to be distinct, and denote by A the set of these distinct elements, then the mapping $g : A \to \mathbb{Q}_+$ defined by $g(x) = [x]$, where $[x]$ is the equivalence class of x in $\mathbb{Q}_+$, is a surjection. Now we define the mapping $h : \mathbb{Z}_+ \to A$ by $h(1) = \frac{1}{1}$, $h(2) = \frac{2}{1}$, $h(3) = \frac{1}{2}$, $h(4) = \frac{1}{5}$, $\ldots$, that is, we move in the direction of arrows as shown in the above arrangement to find the images of successive element in $\mathbb{Z}_+$. The mapping h is obviously a bijection. Now the mapping $f = g \circ h$ is the required surjection. This proves that $\mathbb{Q}_+$ is denumerable.

In view of Theorem 1.10.3, the following corollary is immediate.

Corollary 1.10.5 Every subset of a countable set is countable.

Proof. Let B be a subset of a countable set A. If B is finite, then it is countable. If B is infinite, then the countability of B follows from Theorem 1.10.3.

Corollary 1.10.6 Let A and B be two sets. (1) If A is countable, and $f : A \to B$ is a surjection, then B is countable. (2) If A is countable, and $f : B \to A$ is an injection, then B is countable.

Theorem 1.10.7 The set $\mathbb{Z}_+ \times \mathbb{Z}_+$ is countable.

Proof. The mapping $f : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{Z}_+$ defined by

$$f(i, j) = 2^i.3^j \quad \text{for all } i, j \in \mathbb{Z}_+$$

is injective. Now apply Theorem 1.10.3.

An indexed family of sets $\{A_i\}_{i\in J}$ is said to be a countable family if the index set J is a countable set. By a countable union we mean the union of sets in a countable family.

Theorem 1.10.8 A countable union of countable sets is countable.

Proof. Let $\{A_i\}_{i\in J}$ be a countable family of countable sets. If J is empty, then $\bigcup_{i\in J} A_i$ is empty, and so it is countable. So we assume that J is non-empty. Also without loss of generality, we may assume that $A_i, i \in J$, is non-empty. Since each A_i in the family is countable, there exists a bijective mapping $f_i : \mathbb{Z}_+ \to A_i$. We define a surjective mapping $g : \mathbb{Z}_+ \to J$. This is possible since J is countable. Now we define the mapping $h : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \bigcup_{i\in J} A_i$ by $h(n, m) = f_{g(n)}(m)$. To complete the proof, it is enough to show that h is surjective. Let x be any element in $\bigcup_{i\in J} A_i$. Then x is in some A_i. Therefore, there is some m such that $f_i(m) = x$. Since $g : \mathbb{Z}_+ \to J$ is surjective, and $i \in J$, there exists an n in $\mathbb{Z}_+$ such that $g(n) = i$. Then $h(n, m) = f_{g(n)}(m) = f_i(m) = x$.

Example 1.10.9 The set of rational numbers $\mathbb{Q}$ is denumerable. Denoting by $\mathbb{Q}_+$ and $\mathbb{Q}_-$ the set of positive rational numbers and negative rational numbers, respectively, $\mathbb{Q} = \mathbb{Q}_+ \cup \{0\} \cup \mathbb{Q}_-$. It is clear that there is a bijective correspondence between $\mathbb{Q}_+$ and $\mathbb{Q}_-$. Therefore, to show that $\mathbb{Q}$ is denumerable, it is enough to show that $\mathbb{Q}_+$ is denumerable. Now the mapping defined by $f\left(\frac{p}{q}\right) = (p, q)$ is injective. Therefore, $\mathbb{Q}_+$ is countable by corollary 1.10.6. Since $\mathbb{Q}_+$ is a super set of $\mathbb{Z}_+$, and $\mathbb{Z}_+$ is infinite, $\mathbb{Q}_+$ is an infinite set (see Corollary 1.11.13). Therefore, $\mathbb{Q}_+$ is denumerable.

Theorem 1.10.10 A finite Cartesian product of countable sets is countable.

Proof. Let $A_1, A_2, \ldots, A_n$ be n countable sets. We show the result by induction. First we show that $A_1 \times A_2$ is countable. If A_1 or A_2 is empty, then $A_1 \times A_2$ is empty, and so $A_1 \times A_2$ is countable. Otherwise, there are surjections $f : \mathbb{Z}_+ \to A_1$ and $g : \mathbb{Z}_+ \to A_2$. Then the mapping $h : \mathbb{Z}_+ \times \mathbb{Z}_+ \to A_1 \times A_2$ defined by $h(n, m) = (f(n), f(m))$ is surjective, and so $A_1 \times A_2$ is countable. Now, assume that $A_1 \times A_2 \times \ldots \times A_{n-1}$ is countable. The mapping

$$k : A_1 \times A_2 \times \ldots \times A_n \to (A_1 \times A_2 \times \ldots \times A_{n-1}) \times A_n$$

defined by $k(x_1, x_2, \ldots, x_n) = ((x_1 \times x_2, \ldots, x_{n-1}), x_n)$ is a bijection. Since $A_1 \times A_2 \times \ldots \times A_{n-1}$ is countable, and A_n is countable $(A_1 \times A_2 \times \ldots \times A_{n-1}) \times A_n$ is countable as in the initial step. The Theorem now follows.

The Cartesian product of an arbitrary family of countable sets may not be a countable set as we show next.

Definition 1.10.11 A set is said to be uncountable if it is not countable.

Theorem 1.10.12 If $X = \{0, 1\}$, then the set X^ω is uncountable.

Proof. We show that there is no surjective mapping $f : \mathbb{Z}_+ \to X^\omega$. Suppose to the contrary that $f : \mathbb{Z}_+ \to X^\omega$ is a surjection. Then for each n, $f(n)$ is a sequence of zeros and ones. Assume that

$$f(n) = \langle x_{n1}, x_{n2}, \ldots \rangle$$

where each x_{ni}, $i = 1, 2, \ldots$ is either 0 or 1. We define a sequence $\langle y_n \rangle$ in X^ω given by

$$y_n = \begin{cases} 0 & \text{if } x_{nn} = 1, \\ 1 & \text{if } x_{nn} = 0. \end{cases}$$

Then the sequence $\langle y_n \rangle$ is in X^ω, and it is distinct from each $f(n)$; $\langle y_n \rangle$ differs from $f(n)$ at nth place. Therefore, f is not surjective, a contradiction.

Theorem 1.10.13 The power set $\mathbb{P}(\mathbb{Z}_+)$ of $\mathbb{Z}_+$ is uncountable.

Proof. We show that there is no surjective mapping $f : A \to \mathbb{P}(A)$ for any set A. Let $f : A \to \mathbb{P}(A)$ be any mapping. Then $f(a)$ is a subset of A which may or may not contain a. We define $B = \{a \in A : a \notin f(a)\}$. Now let a_0 be an element in A such that $f(a_0) = B$. Then $a_0 \in B$ if and only if $a_0 \in f(a_0)$ if and only if $a_0 \notin B$. This is a contradiction.

Theorem 1.10.14 The interval $]0, 1[$ of real numbers is uncountable.

Proof. Assume that $]0, 1[$ is countable. Since $]0, 1[$ is an infinite set, it is denumerable. Therefore, there is a bijective mapping $f : \mathbb{Z}_+ \to]0, 1[$. We write the decimal expansion for each $f(n)$. If the decimal expansion of a real number in $]0, 1[$ is terminating, then we decrease the last digit by 1 and replace the following zeros by 9's. For instance, the number $\frac{1}{4} = .25000$ is written as $.24999\ldots$. Now we list all the images

$$\begin{aligned} f(1) &= .a_{11}\ a_{12}\ a_{13}\ \ldots \\ f(2) &= .a_{21}\ a_{22}\ a_{23}\ \ldots \\ f(3) &= .a_{31}\ a_{32}\ a_{33}\ \ldots \\ &\vdots \end{aligned} \qquad (*)$$

where each $a_{ij} \in \{0, 1, 2, \ldots, 9\}$. Now we construct a real number z in $]0, 1[$ which is distinct from each $f(x)$. Let $z = x_1 x_2 \ldots$, where $x_1 = 9$ if $a_{11} \neq 9$, and $x_1 = 0$ if $a_{11} = 9$; $x_2 = 9$ if $a_{22} \neq 9$ and $x_2 = 0$ if $a_{22} = 9$ and so on, in general $x_n = 9$ if $a_{nn} \neq 9$ and $x_n = 0$ if $a_{nn} = 9$. Then z differs from $f(n)$ at the nth place for all n. Therefore, $f : \mathbb{Z}_+ \to]0, 1[$ is not a surjection.

We remark that the method of proof used in Theorem 1.10.14 is known as Cantor's diagonal method because the element z is constructed using the digits $a_{11}, a_{22}, a_{33}, \ldots$ on the principal diagonal of the table (*).

Notice that the above proof uses infinite decimal representation, and a good deal of preliminaries are required to arrive at such a representation which depends on algebraic operations. However, the proof of uncountability of $]0, 1[$ or of $\mathbb{R}$ does not depend on a

decimal representation of a real number, or on any other kind of algebraic properties of $\mathbb{R}$. In Chapter 6, a proof of uncountablity of $\mathbb{R}$ is given that requires only order properties of $\mathbb{R}$.

Corollary 1.10.15 The set $\mathbb{R}$ of real numbers is uncountable.

Proof. By Corollary 1.10.5, if $\mathbb{R}$ were a countable set, the $]0,1[$ would be a countable set. This contradicts Theorem 1.10.14. Therefore, $\mathbb{R}$ is uncountable.

Example 1.10.16. The set of irrational numbers is uncountable. The set of irrational numbers is $\mathbb{R} - \mathbb{Q}$. If $\mathbb{R} - \mathbb{Q}$ were a countable set, then $\mathbb{R} = (\mathbb{R} - \mathbb{Q}) \cup \mathbb{Q}$ would be a countable set, a contradiction.

EXERCISE

1. Show the following:
   (a) $]0,1[$ and $]-1,1[$ are equinumerous.
   (b) $]-1,1[$ is equinumerous with $\mathbb{R}$.
2. Let $\mathcal{F}$ be the collection of all finite subsets of A. If A is infinite, then show that A and $\mathcal{F}$ are equinumerous.
3. Show that if $A \times B$ is finite, then A and B are finite.
4. Prove that if A is denumerable set, and B is a finite subset of A, then $A - B$ is denumerable.
5. If A is a denumerable set, and B is a finite set, then $A \cup B$ is denumerable.
6. Show that every infinite subset of a denumerable set is denumerable.
7. Let $X = \{0,1\}$, and $\mathbb{B}$ be the set of all countable subsets of X^ω. Show that X^ω and $\mathbb{B}$ are equinumerous.
8. Let A be any non-empty set, and 2^A is the set of all mappings from A into $\{0,1\}$. Prove that the power set $\mathbb{P}(A)$ and 2^A are equinumerous.
9. Prove that if A and B are denumerable sets, then $A \times B$ is denumerable.
10. Show that the sets $\mathbb{Z}$, $\mathbb{Z} \times \mathbb{N}$, $\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{Q} \times \mathbb{Q}$ are denumerable sets.
11. Show that if A and B are equinumerous sets, and A is denumerable, then B is denumerable.
12. Show that any superset of a nondenumerable infinite set is nondenumerable.
13. Show that the set of all irrational numbers between 0 and 1 is not denumerable.
14. Let A be a subset of positive real numbers. Let b be a bound such that the sum of elements in any finite subset of A is less than b. Show that A is countable.

15. (i) A real number is said to be algebraic (over the rationals) if it satisfies a polynomial equation
$$a_0x^n + a_1x^{n-1} + \cdots + a_n = 0 \quad (n \geq 1, a_0 \neq 0)$$
where $a_0, a_1, \ldots, a_n$ are rational numbers. Show that the set of algebraic numbers is countable. (Assume that every polynomial equation has finite number of roots.)

    (ii) A real number is said to be transcendental if it is not algebraic. Show that the set of all transcendental numbers is uncountable (It is somewhat surprizing that only two transcendental number are known to us so far. These are e and π).

16. (i) Let $\mathcal{A}$ be a collection of circular disks in the plane, no two of which intersect. Show that $\mathcal{A}$ is countable.

    (ii) Show that the collection of non-intersecting circles in the plane is not countable.

    (iii) Is the collection of figure eights in the plane, no two of which intersect countable?

1.11 Axiom of Choice, Well-ordered Sets and Infinite Sets

In this section, it is established that every set can be well-ordered, that is, on every set we can define an order which is a well-order on the set. Also, it is shown that an infinite set is equinumerous to a proper subset of itself. For both the purposes, however, we need the following axiom.

Axiom of Choice. Let $\mathcal{A}$ be a collection of disjoint non-empty sets. Then there exists a set C having exactly one element in common with each set in $\mathcal{A}$, that is, for each $A \in \mathcal{A}$, $C \cap A$ is singleton.

It may be noted that the axiom of choice has generated a great deal of controversy in the past. There were many mathematicians who were not willing to accept this axiom. There are many theorems in mathematics which have been proved using the axiom of choice. Two such theorems, mentioned above, will, of course, be proved in this section. We prepare the necessary background.

Lemma 1.11.1 Given a collection $\mathcal{A}$ of non-empty sets (not necessarily disjoint), there exists a mapping $\phi : \mathcal{A} \to \bigcup\{A : A \in \mathcal{A}\}$ satisfying the condition that $\phi(A) \in A$ for each set A in $\mathcal{A}$.

(The mapping ϕ is usually called a choice function for the collection $\mathcal{A}$.)

Proof. Let A be an element in $\mathcal{A}$. Define $A' = \{(A, x) : x \in A\}$. Since A is non-empty, A' is non-empty subset of the Cartesian product $\mathcal{A}' \times \bigcup\{A : A \in \mathcal{A}\}$. It is clear that if A_1 and A_2 are distinct elements of $\mathcal{A}$, then A_1' and A_2' are disjoint sets since typical element of A_1' and A_2' are (A_1, x_1) and (A_2, x_2), respectively, and $A_1 \neq A_2$. Then the collection
$$\mathbb{B} = \{A' : A \in \mathcal{A}\}$$
is a collection of disjoint non-empty subsets of $\mathcal{A} \times \bigcup\{A : A \in \mathcal{A}\}$. By the axiom of choice, there exists a set C having exactly one element in common with each $A' \in \mathbb{B}$. We show

that the set C gives the required choice function. Note that $C \subseteq \mathcal{A} \times \bigcup\{A : A \in \mathcal{A}\}$, and that for each $A \in \mathcal{A}$, C contains exactly one pair (A, x) from A', where $x \in A$. Now setting $\phi(A) = x$ if $(A, x) \in C$. Then $\phi(A) \in A$, and ϕ is our required choice function.

Lemma 1.11.2 Let X be a non-empty set. By Lemma 1.11.1, there exists a choice function ψ such that $\psi(B) \in B$ for every non-empty subset B of X. We put $B^* = B - \psi(B)$, $\emptyset^* = \emptyset$, where $\emptyset$ denotes the empty set. Then

$$B^* \subseteq C \subseteq B \quad \text{implies that } B^* = C \quad \text{or } C = B$$

for all $B, C \subseteq X$.

Proof. Since B^* is smaller than B by one point $\psi(B)$ if B is a non-empty subset of X, and $B^* = B$ if B is empty, the lemma follows.

Lemma 1.11.3 Let $\mathcal{A}$ be a collection of subsets of a set X such that

(i) $X \in \mathcal{A}$,

(ii) If $\{A_\gamma\}_{\gamma \in \Gamma}$ is a sub collection of $\mathcal{A}$, then $\bigcap\{A_\gamma : \gamma \in \Gamma\}$ is in $\mathcal{A}$,

(iii) $A \in \mathcal{A}$ implies that $A^* \in \mathcal{A}$.

If $\mathcal{A}_0$ is the intersection of all collections satisfying (i)–(iii), then $\mathcal{A}_0$ also satisfies the properties (i)–(iii).

Proof. The easy proof is left to the reader.

We introduce the order $\leq$ in the set $\mathcal{A}_0$ of Lemma 1.11.3 as follows: for $A, B \in \mathcal{A}_0$, $A \leq B$ if and only if $B \subseteq A$. With respect to this order, $\mathcal{A}_0$ is a totally ordered set.

Lemma 1.11.4 $(\mathcal{A}_0, \leq)$ is a totally ordered set.

Proof. Let $\mathbb{B}_0$ be the set of elements of $\mathcal{A}_0$ which are comparable with every element of $\mathcal{A}_0$. Also let $\mathbb{D}$ be the set of elements B in $\mathcal{A}_0$ such that for any $C \in \mathbb{B}_0$, $C \subseteq B$ or $C^* \supseteq B$, where $C^* = C - \psi(C)$ as defined in Lemma 1.11.2. We show that $\mathbb{D} = \mathcal{A}_0$. To show this, it is enough to show that $\mathbb{D}$ satisfies the conditions (i)–(iii) of Lemma 1.11.3. Since $C \subseteq X$ for every C in $\mathbb{B}_0$, so the condition (i) is satisfied by $\mathbb{D}$. To verify the condition (ii), assume that $B_\gamma \in \mathbb{D}$, $\gamma \in \Gamma$ and $C \in \mathbb{B}_0$. Then for each $\gamma \in \Gamma$, either $C \subseteq B_\gamma$ or $C^* \supseteq B_\gamma$. If $C \subseteq B_\gamma$ for each $\gamma \in \Gamma$, then

$$C \subseteq \bigcap\{B_\gamma : \gamma \in \Gamma\}.$$

If $C^* \supseteq B$ for some $\gamma \in \Gamma$, then

$$C^* \supseteq \bigcap\{B_\gamma : \gamma \in \Gamma\}.$$

Therefore, $\bigcap\{B_\gamma : \gamma \in \Gamma\} \in \mathbb{D}$. Now let $B \in \mathbb{D}$ and $C \in \mathbb{B}_0$. If $C^* \supseteq B$, or $B = C$, then $C^* \supseteq B^*$ since $B \supseteq B^*$. If $B \supsetneqq C$, then $C \supseteq B^*$, or $B^* \supseteq C$ since C is comparable with every element of $\mathcal{A}_0$, and $B^* \in \mathcal{A}_0$. In the first case, it follows from Lemma 1.11.2 that $C = B^*$. Thus, in either case, it follows that $B^* \supseteq C$. Therefore, $B^* \in \mathbb{D}$. Thus, $\mathbb{D} = \mathcal{A}_0$.

To show that $\mathcal{A}_0$ is totally ordered, we show that $\mathcal{A}_0 = \mathbb{B}_0$. This is also shown similarly as above by showing that $\mathbb{B}_0$ satisfies the conditions (i)–(iii) of Lemma 1.11.3. $X \in \mathbb{B}_0$ since it is comparable with every element in $\mathcal{A}_0$. To see the condition (ii), let $C_\gamma \in \mathbb{B}_0$, $\gamma \in \Gamma$, and $B \in \mathcal{A}_0$. If $C_\gamma \supseteq B$ for all $\gamma \in \Gamma$, then

$$\bigcap\{C_\gamma : \gamma \in \Gamma\} \supseteq B.$$

If $C_\gamma \subseteq B$ for some $\gamma \in \Gamma$, then

$$\bigcap\{C_\gamma : \gamma \in \Gamma\} \subseteq B.$$

Therefore, $\bigcap\{C_\gamma; \gamma \in \Gamma\}$ is in $\mathbb{B}_0$. Now we verify the condition (iii). Let $C \in \mathbb{B}_0$, and $B \in \mathcal{A}_0$. Since $\mathbb{D} = \mathcal{A}_0$, $B \in \mathbb{D}$, then either

$$C \subseteq B \quad \text{or} \quad C^* \supseteq B.$$

In the first case, $C^* \subseteq B$ holds. Therefore, in either case, C^* and B are comparable. Thus, $C^* \in \mathbb{B}_0$. This proves that $\mathcal{A}_0 = \mathbb{B}_0$, and the proof is complete.

Lemma 1.11.5 $(\mathcal{A}_0, \leq)$ is a well-ordered set.

Proof. Let $\mathbb{B}$ be a non-empty subset of $\mathcal{A}_0$. If $X \in \mathbb{B}$, then X is the first element of $\mathbb{B}$. If X is not in $\mathbb{B}$, then we define:

$$\mathbb{C} = \{B : B \in \mathcal{A}_0 - \mathbb{B} \text{ and } B \text{ is a lower bound of } \mathbb{B}\}.$$

Since $X \in \mathbb{C}$, $\mathbb{C}$ is non-empty. Let $B_0 = \bigcap\{B : B \in \mathbb{C}\}$. Then $B_0 \in \mathbb{B}$, or $\mathbb{C}$. If $B_0 \in \mathbb{B}$, then, by the definition of B_0, B_0 is the first element of $\mathbb{B}$. If $B_0 \in \mathbb{C}$, then $B_0 \supsetneqq B$ for every $B \in \mathbb{B}$. Since $\mathcal{A}_0$ is totally-ordered, B_0^* is comparable with every element of $\mathbb{B}$. Thus,

$$B_0^* \subseteq B, \quad \text{or} \quad B_0^* \supseteq B \quad \text{for every } B \in \mathbb{B}.$$

In the first case, we obtain $B_0 \supset B \supseteq B_0^*$. Therefore, $B = B_0^*$ by Lemma 1.11.2. Thus, in any case $B_0^* \supseteq B$, that is, $B_0^* \leq B$ for every B in $\mathbb{B}$. Therefore, B_0^* is a lower bound of $\mathbb{B}$. Since B_0^* is in $\mathbb{B}$ by the definition of B_0, B_0^* is the first element of $\mathbb{B}$. Thus, $\mathbb{B}$ has a first element, and the proof is complete.

Theorem 1.11.6 We can introduce into every set X such an appropriate order that X becomes a well-ordered set.

Proof. Let $\mathcal{A}_0$ be the collection of subset of X as defined in Lemma 1.11.3. By Lemma 1.11.5, $\mathcal{A}_0$ is a well-ordered set. Let $f : X \to \mathcal{A}_0$ be a mapping defined by

$$f(a) = \bigcap\{A : a \in A, \text{ and } A \in \mathcal{A}_0\}$$

for all $a \in X$. By the condition (i) and (ii) in Lemma 1.11.3, it follows that f is actually a mapping from X into $\mathcal{A}_0$. We denote $f(a)$ by B_a, that is, $f(a) = B_a$. Since $a \in B_a$, $B_a \neq \emptyset$. Thus, $B_a^* \subsetneqq B_a$. Since $B_a^* \in \mathcal{A}_0$, it follows from the definition of $f(a)$ that $a \notin B_a^*$. Since $B_a - B_a^* = \psi(B_a)$, it means that $\psi(B_a) = a$. Hence, if $a \neq a'$, then $\psi(B_a) \neq \psi(B_{a'})$ and it means that $B_a \neq B_a'$. Therefore, f is injective. Now we define order in X as follows:

$$a \leq a' \quad \text{if} \quad f(a) \leq f(a')$$

With this order X is a partially ordered set isomorphic to $f(X)$. Since $\mathcal{A}_0$ is well-ordered, $f(X)$ is a well-ordered set. Therefore, X is well-ordered with respect to the order introduced.

The above theorem was proved by Zermelo in 1904.

Corollary 1.11.7 There exists an uncountable well-ordered set.

The next result shows the existence of a uncountable well-ordered set whose every proper subset is countable. For this we need the following definition:

Definition 1.11.8 Let $(X, \leq)$ be a partially ordered set. Let $a \in X$. Then the set $W(a) = \{x : x \in X \text{ and } x < a\}$ is called a cut of X by a.

Theorem 1.11.9 There exists an uncountable well-ordered set X such that every cut of X is countable.

Proof. By Corollary 1.11.7, there exists an uncountable well-ordered set $(X, \leq)$. From this, it follows that there exists a well-ordered set Y such that at least one cut of Y is uncountable. The set $\{1, 2\} \times X$ in the dictionary order defined by

$$(a_1, x_1) \leq (a_2, x_2)$$

if $a_1 < a_2$, or if $a_1 = a_2$, then $x_1 \leq x_2$, is one such set. Its cut by an element of the form $(2, x)$ is uncountable. Let Y_0 be the set of all those elements a of Y such that the cut $W(a)$ is uncountable. Then Y_0 will have the least element since Y_0 is a subset of the well-ordered set Y. Let Ω be the least element in Y_0. Then $W(\Omega)$ is uncountable, and every cut of $W(\Omega)$ is countable.

The well-ordered set $W(\Omega)$ of the above theorem is called a minimal well-ordered uncountable set.

Corollary 1.11.10 Every countable subset of $W(\Omega)$ has an upper bound in $W(\Omega)$.

Proof. Let A be a countable subset of $W(\Omega)$. For each $a \in A$, the cut $W(a)$ is countable. Therefore, the union $B = \bigcup_{a \in A} W(a)$ is countable. Since $W(\Omega)$ is uncountable, $B \neq W(\Omega)$. Let x be a point in $W(\Omega) - B$. Then x is an upper bound of A. For if $x < a$ for some $a \in A$, then $x \in W(a)$, and so $x \in B$, a contradiction.

In the remainder of this section, we shift our discussion to infinite sets. The axiom of choice is needed to prove the following theorem.

Theorem 1.11.11 For any infinite set A, there exists an injective mapping $f : \mathbb{N}_0 \to A$.

Proof. Let A be an infinite set. Let ϕ be a choice function for the set of subsets of A, excluding the empty set. Let $\mathbb{B}$ be the set of all finite subsets of A. Let $h : \mathbb{N}_0 \to \mathbb{B}$ be a mapping defined recursively (see section 1.13) as

$$h(0) = \emptyset,$$
$$h(n^+) = h(n) \bigcup \{\phi(A - h(n))\}.$$

That is, $h(n^+)$ is obtained from $h(n)$ by putting in one more element from $A - h(n)$. The set $h(n)$ is a finite subset of A. $A - h(n)$ is non-empty since A is infinite, and $h(n)$ is finite. Now we define $f : \mathbb{N}_0 \to A$ by setting

$$f(n) = \phi(A - h(n)).$$

Thus, $f(n)$ is the element of A we added to $h(n)$ to obtain $h(n^+)$. We show that f is injective. Let $m < n$. Then $m^+ \leq n$. Therefore,

$$f(m) \in h(m^+) \subseteq h(n).$$

But, $f(n) \notin h(n)$ since

$$f(n) = \phi(A - h(n)) \in A - h(n).$$

Thus, $f(m) \neq f(n)$.

Corollary 1.11.12 A set is infinite if and only if it is equinumerous to a proper subset of itself.

Proof. Let A be an infinite set. By Theorem 1.11.11, there exists an injective mapping $f : \mathbb{N}_0 \to A$. Let g be a mapping on A defined by

$$\begin{aligned} g(f(n)) &= f(n^+) && \text{for } n \in \mathbb{N}_0, \\ g(x) &= x && \text{if } x \notin \text{range } (f). \end{aligned}$$

Then g is a bijective mapping from A onto $A - \{f(0)\}$. Hence, an infinite set is equinumerous to its proper subset. The converse of this is Corollary 1.9.10.

Corollary 1.11.13 Every super set of an infinite set is infinite.

Proof. Let A be an infinite set and B is a superset of A. Then, by Corollary 1.11.12, there exists an injective mapping $f : A \to A$ such that $f(A) \neq A$. Define a mapping $g : B \to B$ by

$$g(x) = \begin{cases} f(x) & \text{if } x \in A, \\ x & \text{if } x \in B - A. \end{cases}$$

Then the mapping $g : B \to B$ is injective, and that $g(B) \neq B$.

Corollary 1.11.14 Every subset of a finite set is finite.

Proof. Let A be a finite set, and B be a subset of A. Suppose B is infinite. By Corollary 1.11.13, A is infinite. This is a contradiction.

Theorem 1.11.15 Let $f : A \to B$ be a bijective correspondence. If A is infinite, then B is infinite (Fig. 1.4).

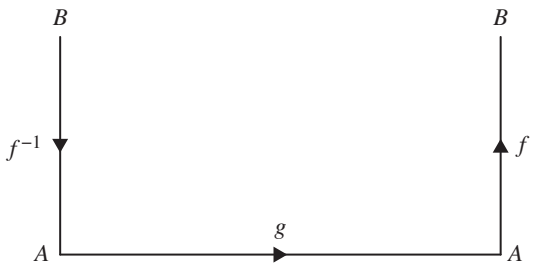


Fig. 1.4

Proof. Since A is infinite, there exists an injective mapping $g : A \to A$ such that $g(A) \neq A$. Since $f : A \to B$ is bijective, the inverse mapping $f^{-1} : B \to A$ is bijective. Consequently, the composition $f \circ g \circ f^{-1} : B \to B$ of injections is injective, and

$$(f \circ g \circ f^{-1})(B) = (f \circ g)(f^{-1}(B)) = (f \circ g)(A) = f(g(A)).$$

Since $g(A) \neq A$, $f(g(A))$ is a proper subset of B. Thus, B is infinite.

Corollary 1.11.16 Let $f : A \to B$ be a bijective correspondence. If A is finite, then B is finite.

Theorem 1.11.17 Let A be an infinite set, and $x_0 \in A$. Then $A - \{x_0\}$ is infinite.

Proof. Since A is infinite, there exists an injective mapping $f : A \to A$ such that $f(A) \neq A$. Now either $x_0 \in f(A)$, or $x_0 \in A - f(A)$. If $x_0 \in f(A)$, then there exists an element $x_1 \in A$ such that $f(x_1) = x_0$. Then the mapping $g : A - \{x_0\} \to A - \{x_0\}$ defined by

$$g(x) = \begin{cases} f(x) & \text{if } x \neq x_1, \\ x_2 & \text{if } x = x_1 \in A - \{x_0\}, \end{cases}$$

where x_2 is an arbitrarily fixed element in the non-empty set $A - f(A)$, is injective. Also, $g(A - \{x_0\}) = f(A - \{x_0, x_1\}) \cup \{x_2\} \neq A - \{x_0\}$. Thus, $A - \{x_0\}$ is infinite in this case. If $x_0 \in A - f(A)$, then the mapping $g : A - \{x_0\} \to A - \{x_0\}$ defined by

$$g(x) = f(x) \quad \text{for all} \quad x \in A - \{x_0\}$$

is injective, and $$g(A - \{x_0\}) = f(A) - \{f(x_0\} \neq A - \{x_0\}.$$

Therefore, in either case, $A - \{x_0\}$ is infinite.

Corollary 1.11.18 Let A be an infinite set, and $x_0, x_1, \ldots, x_k$ be k elements in A. Then $A - \{x_0, x_1, \ldots, x_k\}$ is an infinite set.

Proof. The proof follows by mathematical induction on k and the above theorem.

Recall the definition of a chain in a partially ordered set. A chain A in a partially ordered set P is said to be maximal chain if for all chains B in P containing A it is the case that $A = B$, that is, A is not properly contained in any other chain.

Later, we shall need the following proposition that is equivalent to the axiom of choice, and is known as the maximum principle.

Theorem 1.11.19 (Maximum Principle) Every chain in a partially ordered set is contained in a maximal chain.

EXERCISE

1. Prove that the sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are infinite.
2. Prove that if A is infinite, then $A \times A$ is infinite.
3. Prove that if A is an infinite set, and B is any set, then $A \cup B$ is infinite.
4. Let A and B be sets such that $A \cup B$ is infinite, then either A is infinite or B is infinite.
5. Show that the set $\mathbb{R}$ of real numbers with usual ordering $\leq$ is not a well-ordered set.
6. Prove that for any $n \in \mathbb{N}_0$, all well-ordered sets with n-elements are isomorphic.
7. Show that every well-ordered set A has the least upper bound property, that is every non-empty subset which is bounded from above has its supremum in A.
8. Prove that in a well-ordered set every element except the greatest (if it exists) has its successor.
9. The sets $\{1, 2\} \times \mathbb{Z}_+$ and $\mathbb{Z}_+ \times \{1, 2\}$ are well-ordered in the dictionary order. Are they of same order type?

1.12 Cardinal Numbers

The concept of equinumerosity, introduced in Section 1.9, becomes the basis for defining the concept of a cardinal number. However, before introducing such a concept, we discuss an important theorem that can be used to decide the equinumerosity of two sets.

Theorem 1.12.1 (Schroeder-Bernstein) If X and Y are two sets each of which is equinumerous with a subset of the other, then X and Y are equinumerous.

Proof. Let X be equinumerous with a subset B of Y. Then there exists a bijective mapping $f : X \to B$, so that $f : X \to Y$ is injective. Similarly, there exists an injective mapping $g : Y \to X$. To establish the theorem, we need to show that there is a bijective mapping $F : X \to Y$. If $f : X \to Y$ is surjective, then we can take $F = f$. If $g : Y \to X$ is surjective, we can take $F = g^{-1}$. If both f and g are not surjective, we proceed as follows: Let x be any element in X. If $g^{-1}(x)$ exists, then call it the first ancester of x. Similarly, if $f^{-1}(g^{-1}(x))$ exists, call it the second ancester of x. Again if $g^{-1}(f^{-1}(g^{-1}(x)))$ exists, call it the third ancester of x, and so on, x itself can be called 0th ancester of x. Then there are three possibilities. An element $x \in X$ has infinite number of ancesters, even number of ancesters, or odd number of ancesters. Accordingly, we define the three sets.

$$X_i = \{x \in X : x \text{ has infinite number of ancesters}\},$$
$$X_o = \{x \in X : x \text{ has odd number of ancesters}\},$$
$$X_e = \{x \in X : x \text{ has even number of ancesters}\}.$$

Similarly, we define the three subsets Y_i, Y_o and Y_e of Y. Then it is easily seen that f maps X_i onto Y_i, X_e onto Y_o, and that g^{-1} maps X_o onto Y_e.

Now F is given by (see Fig. 1.5)

$$F(x) = \begin{cases} f(x) & \text{if } x \in X_i \cup X_e, \\ g^{-1}(x) & \text{if } x \in X_o. \end{cases}$$

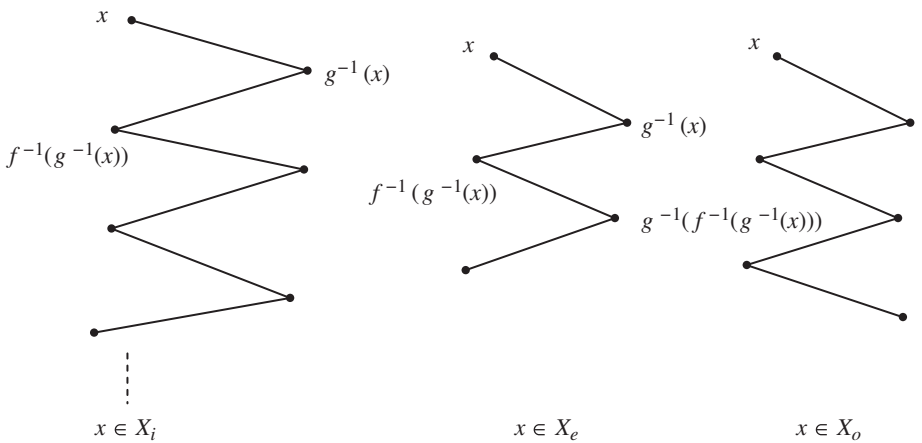


Fig. 1.5

Example 1.12.2 The intervals $[0,1]$ and $]0,1[$ are equinumerous. The mapping $f:[0,1]\to]0,1[$ defined by $f(x)=\sin x$ if $x\in]0,1]$ and $f(0)=\sin x_0$ where x_0 is a point in $]1,\frac{\pi}{2}[$, is injective. Also the identity mapping $i:]0,1[\to[0,1]$ is injective. So, by Schroeder-Bernstein Theorem, the assertion follows.

To every set X a cardinal number denoted by $|X|$ is assigned. The equality $|X|=|Y|$ holds if X and Y are equinumerous.

It is customary to denote the cardinal numbers of the sets $\emptyset$, $\{\emptyset\}$, $\{\emptyset,\{\emptyset\}\}$, $\{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\},\ldots$ by 0, 1, 2, 3, ..., respectively. These are called finite cardinal numbers. Thus, if a set is finite having k elements, its corresponding cardinal number is denoted by k. The cardinal number associated with the set of natural number $\mathbb{N}_0$ is denoted by $\aleph_0$ (aleph zero), and the cardinal number assigned to the set of real numbers $\mathbb{R}$ is denoted by c (continuum). Since $\mathbb{R}$ is equinumerous with any open interval $]a,b[$ and also with any closed interval $[a,b]$ which contains an open interval, these open and closed intervals are also assigned cardinal number c.

Let A and B be two sets. Then

$$|A|\le|B| \qquad \text{if } A \text{ is equinumerous with a subset of } B,$$
$$|A|<|B| \qquad \text{if } |A|\le|B| \text{ and } |A|\ne|B|.$$

From the above definition of order $<$, it is now obvious that

$$0<1<2<\ldots<\aleph_0<c.$$

At this point, we can state one of the most famous unresolved questions in mathematics raised by Cantor: Is there a cardinal number greater than $\aleph_0$ and less than c. No one has answered this question till date. Cantor himself believed that there is no cardinal number between $\aleph_0$ and c, that is, c is the next infinite Cardinal number greater than $\aleph_0$. Most mathematicians also believe that this is true. This guess of Cantor has come to be known as 'Cantor's continuum hypothesis'.

There is another question which we can now consider: Is there any cardinal number strictly greater than c. The answer to this question is 'yes' as the following theorem shows.

Theorem 1.12.3 For any set A,

$$|A|<|\mathbb{P}(A)|.$$

Proof. It is clear that A is equinumerous with a subset of $\mathbb{P}(A)$. Therefore, $|A|\le|\mathbb{P}(A)|$. To show that $\mathbb{P}(A)$ is equinumerous with no subset of A, assume the contrary. Then there exists an injective mapping $f:\mathbb{P}(A)\to A$. We define

$$B_0=\{f(B):B\in\mathbb{P}(A) \text{ and } f(B)\notin B\}.$$

Then $B_0\in\mathbb{P}(A)$. Which element of A is $f(B_0)$? It must be true that $f(B_0)\in B_0$, or $f(B_0)\notin B_0$. But each of these possibilities contradicts the definition of B_0.

Theorem 1.12.4 $|\mathbb{P}(\mathbb{N}_0)|=c$.

Proof. Let $I=[0,1[$. Then it is obvious that $|I|=c$. We define $f:\mathbb{P}(\mathbb{N}_0)\to I$ as follows: Let $A\in\mathbb{P}(\mathbb{N}_0)$. Then $f(A)=x$, where $x=.d_0d_1d_2\ldots$ is defined by the condition that $d_n=3$ or 5 according as n is or is not in A. Then f is injective. Again we define

$g : I \to \mathbb{P}(\mathbb{N}_0)$ as follows: Let $x = .b_0b_1b_2\ldots$ be the binary expansion of a number x in $[0, 1[$. Then $g(x)$ is that subset A of $\mathbb{N}_0$ defined by

$$A = \{n : b_n = 1\}.$$

Then g is injective. Now we conclude the proof by Schroeder-Berstein Theorem.

Example 1.12.5 Let $I = [-1, 1]$. It is obvious that $|I| \leq |\mathbb{R}| = c$. On the other hand since the mapping $f : \mathbb{R} \to [0, 1]$ defined by $f(x) = \left(\frac{2}{\pi}\right) \tan^{-1} x$ is injective, it follows that $c \leq |I|$. Thus, $c = |I|$.

Corollary 1.12.6 $\aleph_0 < c$.

Proof. By Theorem 1.12.3, $|\mathbb{N}_0| < |\mathbb{P}(\mathbb{N}_0)|$. Therefore, by Theorem 1.12.4, $\aleph_0 < c$.

Addition, multiplication and exponents of cardinal numbers are defined as follows: Let A and B be any two disjoint sets. Then

$$\begin{aligned} |A| + |B| &= |A \cup B|, \\ |A| \times |B| &= |A \times B|, \\ |A|^{|B|} &= |A^B|. \quad (A^B \text{ denotes the set all mappings from } B \text{ to } A) \end{aligned}$$

We can also define addition and multiplication of infinite number of cardinal numbers. Let $\{A_\gamma\}_{\gamma\in\Gamma}$ be a family of disjoint sets.

$$\sum |A_\gamma| = \left|\bigcup\{A_\gamma : \gamma \in \Gamma\}\right|,$$

$$\prod_{\gamma\in\Gamma} |A_\gamma| = \left|\prod_{\gamma\in\Gamma} A_\gamma\right|.$$

Example 1.12.7 1. $4 + 3 = 7$.

2. $\aleph_0 + \aleph_0 = \aleph_0$.

Let A and B be the set of even and odd natural numbers, respectively. Then, $|A| = \aleph_0$, $|B| = \aleph_0$. Since $\mathbb{N}_0 = A \cup B$, it follows that $\aleph_0 + \aleph_0 = \aleph_0$.

3. $\aleph_0 + c = c$.

Since the open interval $]0, 1[$ and the set $\mathbb{R}$ of real numbers are equinumerous, $|]0, 1[| = |\mathbb{R}| = c$. Let $S = \mathbb{N}_0 \cup]0, 1[$. Since $\mathbb{N}_0$ and $]0, 1[$ are disjoint, $|S| = |\mathbb{N}_0| + |]0, 1[| = \aleph_0 + c$. On the other hand, since $\mathbb{R} \approx]0, 1[\subseteq S$ and $S \approx S \subseteq \mathbb{R}$, by Schroeder-Bernstein Theorem, $S \approx \mathbb{R}$. Therefore, $\aleph_0 + c = c$.

4. $x + 0 = x$ for any cardinal number x.

For any cardinal number x, there exists a set A such that $|A| = x$. Since $A \cup \emptyset = \emptyset$

$$\begin{aligned} x + 0 &= |A| + |\emptyset| \\ &= |A \cup \emptyset| \\ &= |A| \\ &= x. \end{aligned}$$

5. $c + c = c$.

Let $I =]0,1[$ and $J =]1,2[$. Then $|I| = c$, $|J| = c$. Then $c + c = |I \cup J|$. On the other hand $\mathbb{R} \approx I \subseteq I \cup J$, and $I \cup J \approx I \cup J \subseteq \mathbb{R}$. Therefore, by Schroeder-Bernstein Theorem, $\mathbb{R} \approx I \cup J$. Thus, $c + c = c$.

Example 1.12.8 Let x be any cardinal number. Then

(i) $1x = x$,

(ii) $0x = 0$,

(iii) $\aleph_0\aleph_0 = \aleph_0$,

(iv) $cc = c$.

Solution. Let A be a set such that $|A| = x$.

(i) Since $\{1\} \times A$ is equinumerous with A, $1x = x$.

(ii) $\emptyset \times A = \emptyset$ implies that $0x = 0$.

(iii) Since $\mathbb{N}_0 \times \mathbb{N}_0 \approx \mathbb{N}_0, \aleph_0\aleph_0 = \aleph_0$.

(iv) It is easily seen that $c \leq cc$ so that we prove $cc \leq c$. Suppose that each $x \in]0,1[$ is expressed by its infinite decimal expansion. For instance, $\frac{1}{2}$ is expressed as $.4999\ldots$, instead of .5. In this way, there is a unique expression for each $x \in]0,1[$. It is easily seen that the mapping $f :]0,1[\times]0,1[\to]0,1[$ defined by

$$f(.x_1x_2\ldots, .y_1y_2\ldots) = .x_1y_1x_2y_2\ldots.$$

is injective. Thus, $]0,1[\times]0,1[$ is equinumerous with a subset of $]0,1[$. So that $cc \leq c$.

Example 1.12.9 Let A be a set. Then $2^{|A|} = |\mathbb{P}(A)|$. To see this, let $B = \{0,1\}$. For a subset D of A, let χ_D be the characteristic function of D. Then the mapping from $\mathbb{P}(A)$ to B^A which maps D to χ_D is bijective. Therefore, $|\mathbb{P}(A)| = |B^A| = |B|^{|A|} = 2^{|A|}$.

The following results on exponentiation of cardinal numbers hold. Let a, b, x, y, z be cardinal numbers. Then

(i) $a^x a^y = a^{x+y}$,

(ii) $(z^y)^x = z^{yx}$,

(iii) $(ab)^x = a^x b^x$.

Also, in view of Theorem 1.12.4 and Example 1.12.9, it follows that $2^{\aleph_0} = c$. Also, by Theorem 1.12.3 $c < 2^c < 2^{2^c} < \ldots$, and we have the following order among the infinite sequence of cardinal numbers.

$$0 < 1 < 2 < \ldots < \aleph_0 < c < 2^c < 2^{2^c} < \ldots.$$

The cardinal number of an infinite set is called a transfinite cardinal number. Thus, $\aleph_0$, c, 2^c, ... are all transfinite cardinal numbers. The continuum hypothesis now becomes the particular case of the following statement known as Generalised Continuum Hypothesis: For any transfinite cardinal number a, there is no cardinal number x such that $a < x < 2^a$.

We close this elementary discussion of cardinal numbers by considering the following example.

Example 1.12.10 Let $\{\mu_\gamma : \gamma \in \Gamma\}$ be a set of cardinal numbers. Then $\mu = \sum_{\gamma\in\Gamma} \mu_\gamma$ is obviously a cardinal number, and $\mu_\gamma \leq \mu$ for every $\gamma \in \Gamma$. Then, by Theorem 1.12.3 $\mu < 2^\mu$. Thus, if M is the set of all cardinal numbers, then we should obtain a cardinal number ν which is greater than every cardinal number in M. Therefore, $\nu \notin M$. But we assumed that M contains all cardinal numbers. This example shows that we should not regard M as a set. Formal set theories were developed to avoid such antinomies.

Cantor conjectured that the (Generalised) continuum hypothesis was true. In 1939, Godel proved that the inclusion of generalised continuum hypothesis would not make a consistent set theory inconsistent, that is, if the generalised continuum hypothesis is added to the current axioms of set theory, then any contradiction implied by this system of axioms could be formulated as a contradiction implied by the initial axioms (without the generalised continuum hypothesis) alone. In 1963, Paul Cohen proved that generalised continuum hypothesis could not be proved from other axioms of set theory. The debate is still on whether generalised continuum hypothesis is true or not.

EXERCISE

1. If X and Y are two sets, then show that $X \times Y$ and $Y \times X$ are equinumerous.
2. If $A - B$ and $B - A$ are equinumerous, then show that A and B are equinumerous.
3. Let X, Y, Z and W be sets with $X \cap Z = \emptyset = Y \cap W$, and X is equinumerous with Y, and Z is equinumerous with W. Then $X \cup Z$ and $Y \cup W$ are equinumerous.
4. Let $(A_\gamma)_{\gamma\in\Gamma}$ and $(B_\gamma)_{\gamma\in\Gamma}$ be two families of disjoint sets such that A_γ and B_γ are equinumerous for each $\gamma \in \Gamma$. Then $\bigcup_{\gamma\in\Gamma} A_\gamma$ and $\bigcup_{\gamma\in\Gamma} B_\gamma$ are equinumerous.
5. Let X, Y, Z and W be sets such that X and Y are equinumerous, and Z and W are equinumerous. Then $X \times Z$ and $Y \times W$ are equinumerous.
6. Find a one-to-one correspondence between the open interval $]0, 1[$ and $\mathbb{R}$ that takes rationals to rationals and irrationals to irrationals.
7. Construct a one-to-one correspondence between the closed unit interval $[0, 1]$ and the open interval $]0, 1[$.
8. Show that there is no largest cardinal number.
9. Show that the mapping $f :]0, 1[\times]0, 1[\to]0, 1[$ defined by $f(.x_1x_2\ldots, .y_1y_2\ldots) = .x_1y_1x_2y_2\ldots$ in Example 1.12.8 is bijective.
10. Let x and y be cardinal numbers. Prove that
    (a) if $xy = 0$, then $x = 0$ or $y = 0$.
    (b) if $xy = 1$, then $x = 1$ and $y = 1$.
11. Let x, y, z, a and b be cardinal numbers. Prove that

(a) if $x \leq y$, then $xz \leq yz$.

(b) if $x < y$ and $z \neq 0$ then $xz < yz$.

(c) $n\aleph_0 = \aleph_0$ for any finite cardinal number n.

12. Let x, y, a and b be cardinal numbers. Then prove that

(a) $a^0 = 1$, $a^1 = a$, $1^a = 1$, and $0^a = 0$ if $a \neq 0$.

(b) if $a \leq b$ and $x \leq y$, then $a^x \leq b^y$.

13. (i) $n^{\aleph_0} = c = {\aleph_0}^{\aleph_0}$ for any finite cardinal number n.

(ii) $c^{\aleph_0} = c = c^n$ for any finite cardinal number n.

(iii) $\aleph_0 c = c$.

14. Let $\mathbb{C}$ denote the set of all complex numbers. Prove that $|\mathbb{C}| = c$.

15. Let $H = \{f : \mathbb{R} \to \mathbb{R} \text{ is a real valued function}\}$. Then show that

$$|H|^n = |H|^{\aleph_0} = |H|^c = |H|$$

for all finite cardinal numbers n.

16. Let A be any set having cardinal number k. A permutation of A is a bijective mapping from A onto A. The factorial of k, denoted $k!$ is defined by

$$k! = |\{f : f \text{ is a permutation of } A\}|.$$

Show that $k!$ is well-defined, that is, the value of $k!$ does not depend on the choice of the set A.

17. Show that for an infinite cardinal number k, we have $k! = 2^k$, where $k!$ is defined as in the preceding Exercise 16.

18. If the domain of a mapping is countable, then the range of the mapping is countable.

1.13 Ordinal Numbers

Recall the definition of a totally ordered set and of a well-ordered set. Two totally ordered sets X and Y are said to be of same order type if they are isomorphic. The order type of a well-ordered set is called the ordinal number of the set. The ordinal number of a well-ordered set A may be denoted by ord A.

By Theorem 1.11.6, every set can be well-ordered. The ordinal number of the well-ordered sets

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots$$

are denoted by 0,1,2,3,4,... respectively. These are called finite ordinal numbers. Although the same symbols 0,1,2,3,... are said to denote natural numbers, cardinal numbers, and also ordinal numbers, it is usually clear from the context what these symbols refer to.

The ordinal number of a well-ordered infinite set is called a transfinite ordinal number. The set of natural numbers $\mathbb{N}_0$ with usual order $\leq$ is a well-ordered set. Its ordinal number is denoted by ω.

It is obvious that if any two well-ordered sets X and Y are isomorphic, then X and Y have the same cardinal numbers, that is, $|X| = |Y|$. We say that the ordinal number α of a well-ordered set X has the cardinality $|X|$, or α is of cardinality $|X|$. Thus, the ordinal number 0 is of cardinality 0, the ordinal number 1 is of cardinality 1, and so on. The transfinite ordinal number ω is of cardinality $\aleph_0$, the cardinality of the set of natural numbers $\mathbb{N}_0$.

Are there other ordinal numbers? The following discussion shows that there are many ordinal numbers having the same cardinality.

We define the order relation $<$ between two ordinal numbers α of X and β of Y as follows: $\alpha < \beta$ if X is isomorphic with a subset $Y(b) = \{y : y \in Y \text{ and } y < b\}$ of Y for some element $b \in Y$.

The subset $Y(b)$ of Y is called a cut of Y (see Section 1.11).

In the following, we show that every two ordinal numbers are comparable according to the order $<$.

Lemma 1.13.1 Let A be a well-ordered set, and f an isomorphism of A onto a subset of A. Then $a \leq f(a)$ for every element a of A.

Proof. Assume that $a > f(a)$ for some $a \in A$. Then the set $A' = \{a : a > f(a)\}$ is non-empty subset of A. Since A is well-ordered, there is the first element $a_0 \in A'$. Since $a_0 \in A'$,

$$a_0 > f(a_0).$$

Therefore, $f(a_0) \notin A'$. By the definition of A',

$$f(a_0) \leq f(f(a_0)).$$

Since f is an isomorphism, and $a_0 > f(a_0)$, it follows that

$$f(f(a_0)) < f(a_0).$$

Thus, $f(a_0) < f(a_0)$ which is a contradiction, and the assertion of the lemma follows.

Theorem 1.13.2 A well-ordered set A is isomorphic to no cut of A.

Proof. Assume, on the contrary, that f is an isomorphism of A onto a cut

$$A(b) = \{a : a \in A \text{ and } a < b\}.$$

By Lemma 1.13.1, $b \leq f(b)$. By the definition of f, $f(b) \in A(b)$, and so $f(b) < b$. This is a contradiction.

If α and β are two ordinal numbers, by the above theorem, it follows that the statements $\alpha = \beta$, and $\alpha < \beta$ are not both true simultaneously. The following theorem does a little more.

Theorem 1.13.3 Let α and β be two ordinal numbers. Then no two of the following three cases are true at the same time:

$$\alpha < \beta, \qquad \alpha = \beta, \qquad \alpha > \beta.$$

Proof. Let A and B be two well-ordered sets with ordinal numbers α and β, respectively. We assume that $\alpha < \beta$, and $\beta < \alpha$. Then A is isomorphic with a cut of B, and B is isomorphic with a cut of A. It follows that A would be isomorphic with a cut of A. This contradicts Theorem 1.13.2. The other cases follow similarly.

Example 1.13.4 The well-ordered set $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$ is isomorphic with a cut of the well-ordered set $A_0 = \{0, 1, 2, \ldots, 0^*\}(n < 0^*)$ namely $A(0^*) = \{0, 1, 2, \ldots\} = \mathbb{N}_0$. The set A_0 is isomorphic with a cut of the well-ordered set $A_1 = \{0, 1, 2, \ldots, 0^*, 1^*\}$, namely $A(1^*) = \{0, 1, 2, \ldots, 0^*\}$, and so on. The ordinal number of the well-ordered set $\{1, 3, 5, \ldots, 2, 4, 6, \ldots\}$ is greater than ω.

Theorem 1.13.5 Let τ be an ordinal number, and $W(\tau)$ be the set of ordinal numbers less than τ and not less than 0. For $\alpha, \beta \in W(\tau)$, define $\alpha \leq \beta$ if $\alpha < \beta$ or $\alpha = \beta$. Then with respect to this order $\leq$, $W(\tau)$ is a well-ordered set.

Proof. If $\tau = 0$, then $W(\tau) = \emptyset$, and the assertion follows. So we assume that $\tau > 0$. By Theorem 1.13.3, it follows that $W(\tau)$ is a totally ordered set. Let τ be the ordinal number of a well-ordered set T. Let $\alpha \in W(\tau)$. Then $\alpha < \tau$. Therefore, α is the ordinal number of a cut $T(a) = \{t : t \in T \text{ and } t < a\}$ for some $a \in T$. We define $f(\alpha) = a$. It is easily verified that f is an isomorphism between $W(\tau)$ and T. Therefore, $W(\tau)$ is well-ordered with ordinal number τ.

Theorem 1.13.6 Let α and β be two ordinal numbers. Then one and only one of the following three statements is true:

$$\alpha < \beta, \qquad \alpha = \beta, \qquad \alpha < \beta.$$

Proof. By Theorem 1.13.3, at most one of the three cases is true. Thus, it remains to show that at least one of them is true. With notations of Theorem 1.13.5, let

$$W(\alpha) \cap W(\beta) = C.$$

Since the set C is contained in $W(\alpha)$, it is well-ordered. Let γ be the ordinal number of C. We show that $\gamma \leq \alpha$, and $\gamma \leq \beta$. If $W(\alpha) = C$, the ordinal number of C is α since the ordinal number of $W(\alpha)$ is α by Theorem 1.13.5. In the case, $C \subset W(\alpha)$, we denote by δ the first element, which is an ordinal number, of $W(\alpha) - C$. Suppose that η is an ordinal number such that $\eta < \delta$. Then $\eta \in W(\alpha)$ since $\eta < \delta < \alpha$. Furthermore from the definition of δ, it follows that $\eta \in C$. Conversely, let $\eta \in C$. Since $W(\alpha)$ is a well-ordered set by Theorem 1.13.5, we have $\eta < \delta$, or $\eta = \delta$, or $\eta > \delta$. But $\eta \geq \delta$ together with the fact that $\eta \in C \subseteq W(\beta)$ implies that $\delta \leq \eta < \beta$. Therefore, $\delta \in W(\beta)$, and so $\delta \in C$ which contradicts the definition of δ. Therefore, $\eta < \delta$. Now it follows that $C = W(\delta)$, and hence C is a cut of $W(\alpha)$. Therefore, $\gamma < \alpha$. Thus, we have proved that $\gamma \leq \alpha$. The proof $\gamma \leq \beta$ is similar. Now there are four cases:

$\gamma < \alpha$ and $\gamma < \beta$,
$\gamma = \alpha$ and $\gamma = \beta$ which implies $\alpha = \beta$,
$\gamma = \alpha$ and $\gamma < \beta$ which implies $\alpha < \beta$,
$\gamma < \alpha$ and $\gamma = \beta$ which implies $\alpha > \beta$.

In the first case, $\gamma \in C = W(\alpha) \cap W(\beta)$ which implies that $W(\gamma)$ is a cut of C. Since γ is the ordinal number of both C and $W(\gamma)$, this means that $\gamma < \gamma$ contradicting Theorem 1.13.3. Therefore, first case can not happen. The proof is now complete.

Corollary 1.13.7 Every set A of ordinal numbers is well-ordered.

Proof. By Theorem 1.13.6, A is a totally ordered set. Let B be a non-empty subset of A. Let α be an element of B. If α is not the first element of B, let $W(\alpha)$ be the set of ordinal numbers less than α. Let $W(\alpha) \cap B = C$. Since $W(\alpha)$ is a well-ordered set by Theorem 1.13.5, C is a well-ordered set. Therefore, C contains the first element β. Let γ be an element of B. Then, by Theorem 1.13.6, $\beta \leq \gamma$, or $\gamma < \beta$. But $\gamma < \beta$ implies that $\gamma < \beta < \alpha$. Thus, $\gamma \in W(\alpha)$, that is, $\gamma \in C$. This contradicts the fact that β is the first element of C. Therefore, $\beta \leq \gamma$ which means that β is the first element of B.

Corollary 1.13.8 Let A be a well-ordered set, and $B \subseteq A$. Let α and β be the ordinal numbers of A and B, respectively. Then $\beta \leq \alpha$.

Proof. Suppose that $\alpha < \beta$. Then A is isomorphic with a cut $B(a) = \{b \in B : b < a\}$ for some $a \in B$. Let f be this isomorphism of A onto the cut $B(a)$. Since $B(a) \subset B \subseteq A$, f is an isomorphism of A onto its subset $B(a)$. By Theorem 1.13.1, $a \leq f(a)$. On the other hand, $f(a) \in B(a)$ so that $f(a) < a$. But this is a contradiction. Therefore, $\beta \leq \alpha$.

We next discuss the operations of addition, multiplication and exponentiation of ordinal numbers.

Let $(A, \leq)$ and $(B, \leq')$ be two well-ordered sets such that A and B are disjoint. We define the total order relation $\leq^*$ on $A \cup B$ as follows:

(i) $a \leq^* b$ if $a \leq b$ for all $a, b \in A$,

(ii) $a \leq^* b$ if $a \leq' b$ for all $a, b \in B$,

(iii) for all $a \in A$ and $b \in B$, we agree that $a \leq^* b$.

It is easy to see that $(A \cup B, \leq^*)$ is a well-ordered set. With the above notation, we define the addition of ordinal numbers.

Let α and β be the ordinal numbers of the disjoint well-ordered sets $(A, \leq)$ and $(B, \leq')$. The ordinal sum of α and β, denoted by $\alpha + \beta$, is the ordinal number of the well-ordered set $(A \cup B, \leq^*)$.

In the case when well-ordered sets A and B in the above definition, are not disjoint, then their ordinal sum is defined by first converting A and B into disjoint well-ordered sets of the same cardinalities as follows: Form the set $A \times \{0\}$ with the order relation $(a, 0) \leq_0 (b, 0)$ if $a \leq b$. Then $(A \times \{0\}, \leq_0)$ is a well-ordered set. Similarly, form the set $B \times \{1\}$ with the order $(a, 1) \leq_1 (b, 1)$ if $a \leq' b$. Then $(B \times \{1\}, \leq_1\}$ is also a well-ordered set. Then the ordinal sum $\alpha + \beta$ is defined to be the ordinal number of the set $(A \times \{0\} \cup (B \times \{1\}), \leq^*)$.

It can be shown that the definition of the ordinal sum $\alpha + \beta$ is independent of the choice of the representative sets A and B. For this one needs to show the following: Let $(A_1, \leq_1)$, $(A_2, \leq_2)$, $(B_1, \leq_1')$ and $(B_2, \leq_2')$ be well-ordered sets such that $A_1 \approx A_2$, $B_1 \approx B_2$, then $(A_1 \cup B_1, \leq_1^*) \approx (A_2 \cup B_2, \leq_2^*)$. We leave it to the reader to establish this fact.

Example 1.13.9 Since 0 is the ordinal number of the empty set $\emptyset$ and 1 is the ordinal number of the set $\{\emptyset\}$. The ordinal sum $0 + 1 = 1$. In general, $0 + \alpha = \alpha$ for all ordinal numbers α. Similarly, the ordinal sum $5 + 4$ is equal to 9.

Example 1.13.10 We have, $\omega = \operatorname{ord} \mathbb{N}_0$. Let $n = \operatorname{ord}\{a_0, a_1, \ldots, a_{n-1}\}$. Then

$$n + \omega = \operatorname{ord}\{a_0, a_1, \ldots, a_{n-1}, 0, 1, \ldots\} = \omega.$$

$$\omega + n = \text{ord}\{0, 1, 2, \ldots, a_0, a_1, \ldots, a_{n-1}\} > \omega$$

since $\mathbb{N}_0$ is isomorphic with the cut $W(a_0)$ of the set $\{0, 1, 2, \ldots, a_0, a_1, \ldots, a_{n-1}\}$. This shows that the addition of ordinal numbers, is not commutative.

For the purpose of defining the product of ordinal numbers, the dictionary order (see Theorem 1.11.9) is used: Let $(A, \leq_1)$ and $(B, \leq_2)$ be two partially ordered sets. Recall that the dictionary order $\leq$ on $A \times B$ is defined by

$$(a_1, b_1) \leq (a_2, b_2) \text{ if } a_1 <_1 a_2, \text{ or if } a_1 = a_2 \text{ then } b_1 \leq_2 b_2.$$

The dictionary order is easily seen to be a partial order, and if $(A, \leq_1)$ and $(B, \leq_2)$ are well-ordered sets, then the dictionary order is a well-order on $A \times B$.

Theorem 1.13.11 Let $(A, \leq_1)$ and $(B, \leq_2)$ be two well-ordered sets. Then the dictionary order relation $\leq$ is a well-order relation on $A \times B$.

Proof. Since the relation $\leq_1$ and $\leq_2$ are total orders on A and B, respectively, it follows from the definition of $\leq$ that $\leq$ is a total order on $A \times B$. To show that it is a well-order on $A \times B$, let S be a non-empty subset of $A \times B$. Now the projection

$$p_A(S) = \{x : (x, y) \in S \text{ for some } y \in B\}$$

is a non-empty subset of A, and so it contains the least element a. The set

$$\{y \in B : (a, y) \in S\}$$

is non-empty subset of the well-ordered set B. So it contains the least element b. Then it is easily seen that (a, b) is the least element of S. Therefore, $(A \times B, \leq)$ is a well-ordered set.

Theorem 1.13.12 Let $(A_1, \leq_1)$, $(A_2, \leq_2)$, $(B_1, \leq_1')$ and $(B_2, \leq_2')$ be well-ordered sets such that $(A_1, \leq_1) \approx (A_2, \leq_2)$, and $(B_1, \leq_1') \approx (B_2, \leq_2')$. Then

$$(A_1 \times B_1, \leq_1^*) \approx (A_2 \times B_2, \leq_2^*)$$

where $\leq_1^*$ and $\leq_2^*$ are the dictionary orders on $A_1 \times B_1$ and $A_2 \times B_2$, respectively.

Proof. Let $f : A_1 \to A_2$ and $g : B_1 \to B_2$ be isomorphisms. Then the mapping $f \times g : A_1 \times B_1 \to A_2 \times B_2$ defined by $f \times g(a, b) = (f(a), g(b))$ is an isomorphism.

In view of the above theorems, it makes sense to define the product of ordinal numbers as follows:

Let $(A, \leq_1)$ and $(B, \leq_2)$ be well-ordered sets with ordinal numbers α and β, respectively. The ordinal product $\beta\alpha$ of α and β is defined to be the ordinal number of the well-ordered set $(A \times B, \leq)$, where $\leq$ is the dictionary order on $A \times B$.

Example 1.13.13 Let $2 = \text{ord}\{0, 1\}$, and $\omega = \text{ord}\{0, 1, 2, 3, \ldots\}$. Then

$$2\omega = \text{ord}\{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1), \ldots\}.$$

The mapping f defined by

$$f(j, k) = \begin{cases} 2j & \text{if } k = 0, \\ 2j + 1 & \text{if } k = 1 \end{cases}$$

is an isomorphism from $\{0,1,2,\ldots\}\times\{0,1\}$ onto $\{0,1,2,3,\ldots\}$. Therefore, $2\omega=\omega$.

$$\omega 2=\text{ord}\{(0,0),(0,1),(0,2),\ldots,(1,0),(1,1),(1,2)\ldots\}.$$

It can be easily seen that $\omega 2=\omega+\omega$.

An ordinal number α is a limit number if there is no ordinal number preceding it, that is, for every ordinal number $\beta<\alpha$, there exists an ordinal number γ such that $\beta<\gamma<\alpha$.

An ordinal number α is said to be the predecessor of an ordinal number β if $\alpha<\beta$, and there is no ordinal number γ such that $\alpha<\gamma<\beta$. When α is the predecessor of β, β is called the successor of α, and we write $\beta=\alpha+1$; β itself is denoted by α^+. Indeed that the successor ordinal $\alpha^+=\alpha+1$, the addition of the ordinal numbers α and 1, can be easily seen, and so the notation for α^+ is consistent with our definition of ordinal sum. Also, obvious is the fact that every ordinal number has a successor.

In view of Theorem 1.13.5, it is sometimes convenient to identify the ordinal number with the set $W(\tau)=\{\alpha:\alpha<\tau\}$ of ordinal numbers. For example

$$\begin{aligned}
0&\equiv\emptyset\\
1&\equiv\{0\}\\
2&\equiv\{0,1\}\\
3&\equiv\{0,1,2\}\\
&\vdots\\
\omega&\equiv\{0,1,2,3,\ldots\}\\
\omega+1&\equiv\{0,1,2,3,\ldots,\omega\}\\
\omega+2&\equiv\{0,1,2,\ldots,\omega,\omega+1\}\\
&\vdots\\
\omega 2&\equiv\{0,1,2,\ldots,\omega,\omega+1,\ldots\}\\
\omega 2+1&\equiv\{0,1,2,\ldots,\omega,\omega+1,\ldots,\omega 2\}\\
&\vdots\\
\omega 3&\equiv\{0,1,2,\ldots,\omega,\omega+1,\ldots,\omega 2,\omega 2+1,\ldots\}.
\end{aligned}$$

We proceed further as follows

$$\omega 3,\omega 3+1,\ldots,\omega 4,\ldots,\omega 5,\ldots,\omega\omega=\omega^2$$

where ω^2 is the limit number that contains all the preceding ordinal numbers, that is

$$\omega^2=\{0,1,2,\ldots,\omega,\ldots,\omega 2,\ldots,\omega 3,\ldots,\omega 4,\ldots\}.$$

We continue:

$$\omega^2,\omega^2+1,\ldots,\omega^2+\omega,\omega^2+\omega+1,\ldots,\omega^2+\omega 2,\ldots,\omega^2+\omega 3,\ldots,\omega^2+\omega^2=\omega^2 2.$$

Then we have

$$\omega^2 2,\ldots,\omega^2 3,\ldots,\omega^2 4,\ldots,\omega^2\omega=\omega^3.$$

Now we have the powers of ω

$$\omega^3, \ldots, \omega^4, \ldots, \omega^\omega.$$

The limit number ω^ω follows after the limit numbers ω^n. ω^ω is equal to $\bigcup\{\omega^n, n \geq 1\}$, a union of countably many sets. Further, each ω^n is a countable set. Thus, ω^ω is a countable set. Compare this with the result in Theorem 1.10.12 which says that $2^{\aleph_0}$ is uncountable. A careful distinction has to be made between cardinal exponential and an ordinal exponentiation.

Again proceed:

$$\omega^\omega, \omega^\omega + 1, \ldots, (\omega^\omega)^\omega, \ldots, ((\omega^\omega)^\omega)^\omega, \ldots.$$

One can continue. After all these ordinals, we have the limit number ϵ_0.

We again continue:

$$\epsilon_0, \epsilon_0 + 1, \ldots, \epsilon_0 + \omega, \ldots, \epsilon_0 + \omega 2, \ldots, \epsilon_0 + \omega^2, \ldots, \epsilon_0 + \omega^\omega \ldots,$$

$$\epsilon_0 2, \ldots, \epsilon_0 \omega, \ldots, \epsilon_0 \omega^\omega, \ldots, \epsilon_0^2, \ldots, \epsilon_0^3, \ldots, \epsilon_0^\omega, \ldots.$$

Each of the ordinal numbers enumerated so far is still the ordinal number of a countable set.

After the ordinal numbers of countable sets (or countable ordinal numbers) there are ordinal numbers of uncountable sets (or uncountable ordinal numbers). We know that uncountable sets exist, the set of real numbers $\mathbb{R}$ and $\mathbb{P}(\mathbb{N}_0)$ are examples. We do not know any well-ordering of any uncountable set. However, by Theorem 1.11.6, it is known that every set can be given a well-ordering. In particular, $\mathbb{R}$ can be well-ordered, and therefore with this well-ordering, $\mathbb{R}$ will be associated to an ordinal number. There are other ordinal numbers which are equinumerous with $\mathbb{R}$. However, it can be proved that there is the least uncountable ordinal number. To see this, let τ be an uncountable ordinal number, and it is not the least such ordinal number. By Theorem 1.13.5, τ is the ordinal number of the set $W(\tau)$. Since τ is not the least uncountable ordinal number, $W(\tau)$ contains atleast one uncountable ordinal number. Therefore, the set $A = \{\beta : \beta \in W(\tau) \text{ and } \beta \text{ is uncountable}\}$ is non-empty. Since A is a subset of the well-ordered set $W(\tau)$, it contains the least element. Since this least element belongs to A, it is uncountable.

The least uncountable ordinal number is usually denoted by ω_1. Also note that since every ordinal number is the set of all smaller ordinals, ω_1 is the set of all countable ordinal numbers. The cardinality of ω_1, is denoted by $\aleph_1$. The smallest ordinal number of cardinality larger than $\aleph_1$ is denoted by ω_2, and its cardinality is denoted by $\aleph_2$. Continuing in this way corresponding to an ordinal number α, there is a smallest ordinal number ω_α of cardinality $\aleph_\alpha$. Also, it may be noted that for any sequence $\alpha_1, \alpha_2, \ldots$ of ordinal numbers smaller than ω_1, there exists an ordinal number $\alpha < \omega_1$ such that $\alpha_i < \alpha$ for $i = 1, 2, 3, \ldots$.

Note that in Corollary 1.13.7, we have avoided the phrase "the set of all ordinal numbers". As in the case of Russell's paradox, the assumption of the set of all ordinal numbers leads to contradiction.

Theorem 1.13.14 There does not exist a set of all ordinal numbers.

Proof. Let S be the set of all ordinal numbers. Then S is a well-ordered set. The ordinal number σ of S must be the element of S. By Theorem 1.13.5, it follows that

$$\sigma = \operatorname{ord}\{\alpha : \alpha \in S \text{ and } \alpha < \sigma\} < \operatorname{ord} S = \sigma$$

which is a contradiction.

In certain mathematical statements the set of natural numbers acts as an index set. Such statements are sometimes proved using the principle of mathematical induction. Similarly, there are statements in which a given well-ordered set acts as an index set. For instance, if A is a well-ordered set and $\{P_\alpha\}_{\alpha\in A}$ be an indexed family of propositions, then to see that P_α is true for every $\alpha \in A$, we need the following principle known as the principle of transfinite induction.

Theorem 1.13.15 (Principle of Transfinite Induction.) Let A be any well-ordered set, and S be a subset of A such that

(i) $a_0 \in S$, where a_0 is the first element of A,

(ii) for each $a \in A$, the cut $W(a) \subseteq S$ implies that $a \in S$.

Then $S = A$.

The immediate consequence of the above principle is the following theorem which is also known as the principle of transfinite induction in mathematical parlance.

Theorem 1.13.16 Let A be a well ordered set. Suppose P_α denotes a proposition defined for $\alpha \in A$. If

(i) P_{α_0} is true, where α_0 is the first element in A,

(ii) P_α is true whenever P_β is true for every β such that $\alpha_0 < \beta < \alpha$, then P_α is true for every $\alpha \in A$.

Proof. Suppose, if possible, there exists an element $\alpha \in A$ such that P_α is not true. Then the set B of $\alpha \in A$ such that P_α is not true is non-empty, and consequently, it has the first element β_0. Since P_{α_0} is true, $\beta_0 > \alpha_0$. By the definition of β_0, P_β is true for every $\beta < \beta_0$. But then by (ii), P_{β_0} must be true. This is a contradiction. So, by the principle of transfinite induction, it follows that P_α is true for every $\alpha \in A$.

Since the set of natural numbers is a well-ordered set, the principle of mathematical induction becomes a special case of the principle of transfinite induction. The principle of mathematical induction can be used to define a mapping on the set of natural numbers, however, the principle of transfinite induction can not be used to define a mapping on the collection of all ordinal numbers since the collection of all ordinal numbers is not a set.

The elements of a countable set can be written in the form of a sequence. This is however not possible for uncountable sets. For arbitrary well-ordered sets or ordinal numbers the idea of a sequence is extended as follows:

A transfinite sequence in a set X is a mapping from an ordinal numbers α into X. This transfinite sequence is said to be of length α.

Any set can be represented by a transfinite sequence.

Quite often in mathematics, the values of mappings are generated from certain initially defined values using a formula. For example, a very popular sequence of rational numbers $\langle x_n\rangle$ is so defined:

$$x_1 = 1,$$
$$x_{n+1} = \frac{4+3x_n}{3+2x_n} \quad \text{for } n \geq 1.$$

As an another example, in Theorem 1.10.3, we defined a mapping $k : \mathbb{Z}_+ \to C$ as

$$k(1) = \inf C,$$
$$k(n) = \inf\{C - k(\{1, 2, \ldots, n-1\})\} \text{ for all } n \geq 2.$$

Here the value of k at n is defined in terms of the values of k at j, where $j < n$. That is, the value of k at n is constructed in terms of its previously determined values. When a mapping is so defined, we say that it is defined recursively. Specifically, we have the following principle.

Theorem 1.13.17 (Principle of Recursive Definition) Let A be a set, and $a_0 \in A$. For $a \in \mathbb{Z}_+$, let $S_a = \{n : n \in \mathbb{Z}_+, \text{ and } n \leq a\}$. Let

$$\mathcal{F} = \{f : f : S_a \to A \text{ is a mapping for some } a \in \mathbb{Z}_+\}.$$

Let $\phi : \mathcal{F} \to A$ be a mapping. Then there exists a unique mapping $h : \mathbb{Z}_+ \to A$ such that

$$\begin{cases} h(1) = a_0, \\ h(n) = \phi(h|S_{n-1}) \quad \text{for all } n > 1 \end{cases} \tag{$*$}$$

where $h|S_{n-1}$ is the mapping from S_{n-1} into A given by $(h|S_{n-1})(j) = h(j)$ for all $j \in S_{n-1}$.

The above principle of recursive definition determines $h(1)$, and it expresses h at $n > 1$ in terms of the values of h for positive integers less than n. The formula (*) is called a recursion formula or recursion procedure.

One can prove the above theorem rigorously. We accept it without proof.

Example 1.13.18 Let $\langle a_1, a_2, \ldots, \rangle$ be an infinite sequence of real numbers. We define the product $\prod_{k=1}^{n} a_k$ by the equations

$$\prod_{k=1}^{1} a_k = a_1,$$

$$\prod_{k=1}^{n} a_k = \left(\prod_{k=1}^{n-1} a_k\right) a_n \quad \text{for } n > 1.$$

Let A be the set of points of the sequence. The corresponding function ϕ in Theorem 1.13.17 is defined as $\phi(f) = f(m)a_{m+1}$, where $f : S_m \to A$ for some $m \in \mathbb{Z}_+$. Then

$$h(1) = a_1,$$
$$h(n) = \phi(h|S_{n-1}) = h(n-1)a_n.$$

So that the mapping h is given by

$$h(1) = a_1,$$
$$h(n) = (a_1 a_2 \ldots a_{n-1})a_n.$$

Example 1.13.19 Show that there is a unique function $h : \mathbb{Z}_+ \to \mathbb{R}_+$ satisfying the formula

$$h(1) = 3,$$
$$h(n) = (h(n-1) + 1)^{\frac{1}{2}} \quad \text{for } n > 1.$$

For any mapping $f : S_n \to \mathbb{R}_+$, let $\phi(f) = (f(n) + 1)^{\frac{1}{2}}$. Then the mapping $h : \mathbb{Z}_+ \to \mathbb{R}_+$ of Theorem 1.13.17 is

$$\begin{aligned} h(1) &= 3, \\ h(n) &= \phi(h|S_{n-1}) \\ &= [h(n-1) + 1]^{\frac{1}{2}} \end{aligned}$$

and this h is unique.

The following theorem extends the principle of recursive definition to well-ordered sets.

Theorem 1.13.20 (Transfinite Recursion Theorem) Let α be an ordinal number, and A be a set. Let $\mathcal{F}^\alpha$ be the set of all transfinite sequences in A of length less than α, (formally: $\mathcal{F}^\alpha = \{f : f \text{ is a mapping from } \beta \text{ into } A \text{ for some } \beta < \alpha\}$. Let $\phi : \mathcal{F}^\alpha \to A$ be any mapping. Then there exists a unique mapping $h : \alpha \to A$ such that for all ordinal $\gamma < \alpha$

$$h(\gamma) = \phi(h|\gamma)$$

where $h|\gamma$ is the transfinite sequence of all previously defined values of h (and as such a member of $\mathcal{F}^\alpha$).

Corollary 1.13.21 Let $R(f, x)$ be a formula of two variables f and x, where f represents a transfinite sequence and x is a set. For each transfinite sequence f, $R(f, x)$ holds for precisely one set x. Then there is a similar formula which associates, to each ordinal number α a unique set X_α in such a way that for each ordinal number α, the set X_α is the result of applying the formula R to the transfinite sequence $\{X_\beta : \beta < \alpha\}$.

Theorem 1.13.20 provides for the definition by transfinite induction of a mapping whose domain is a given ordinal number given a mapping ϕ. The mapping ϕ represents a construction process by which given a transfinite sequence of length γ, another element of A may be found which is to be the next member of the sequence under construction. Similarly, in Corollary 1.13.21, the rule R corresponds to the construction procedure for finding the next numbers at each stage.

To obtain large ordinal numbers, the operation of exponentiation is defined. In arithmetic of numbers, the addition consists in adding 1 an appropriate number of times ($m + n = m + 1 + 1 + \cdots + 1, 1$ added n times), multiplication consists in adding a number repeatedly ($m \cdot n = m + m + \cdots + m$). In exponentiation, a number is multiplied repeatedly ($m^n = m \cdot m \cdots m$). The same idea works for ordinals.

Let α and β be any ordinals. We define $e_\alpha(\beta)$ recursively by

$$\begin{aligned} e_\alpha(0) &= 1, \\ e_\alpha(\gamma^+) &= e_\alpha(\gamma)\alpha, \\ e_\alpha(\lambda) &= \bigcup\{e_\alpha(\gamma) : 0 < \gamma < \lambda\} \text{ for all limit numbers } \lambda. \end{aligned}$$

The last equality is permitted by Corollary 1.13.21. We write α^β for $e_\alpha(\beta)$.

Example 1.13.22 (i) $\omega^2 = e_\omega(2) = e_\omega(1^+) = e_\omega(1)\omega = e_\omega(0^+)\omega = (e_\omega(0)\omega)\omega = (1\omega)\omega = \omega\omega$.

Similarly, $\alpha^n = \alpha\alpha\cdots\alpha$ (n times) for any ordinal number α.

(ii) $\omega^\omega = e_\omega(\omega) = \bigcup\{e_\alpha(n) : 0 < n < \omega\} = \bigcup\{\alpha^n : 0 < n < \omega\}$.

Therefore, ω^ω is the first ordinal number larger than each ω^n, $n \geq 1$. For various properties of the operation of exponentiation, the reader is referred to the exercises that follow.

The following theorem is the basis for defining the operation of subtraction.

Theorem 1.13.23 Let α and β be any ordinal numbers such that $\alpha \in \beta$. There exists a unique ordinal number δ (their difference) such that $\alpha + \delta = \beta$.

Example 1.13.24 Let $\alpha = 3$, and $\beta = \omega$. Then the difference is ω since $3 + \omega = \omega$. However, note that if $\alpha \in \beta$, then we can not find in general a δ such that $\delta + \alpha = \beta$; for example, there is no such δ that $\delta + 3 = \omega$.

The next theorem gives the equations that characterize addition and multiplication by recursion.

Theorem 1.13.25 Let α and β be any ordinal numbers, and λ be a limit ordinal. Then the following statements hold:

(A1) $\alpha + 0 = \alpha$,

(A2) $\alpha + \beta^+ = (\alpha + \beta)^+$,

(A3) $\alpha + \lambda = \bigcup\{\alpha + \beta : \beta < \lambda\}$,

(M1) $\alpha 0 = \alpha$,

(M2) $\alpha\beta^+ = \alpha\beta + \alpha$,

(M3) $\alpha\lambda = \bigcup\{\alpha\beta : \beta < \lambda\}$.

Theorem 1.13.26 (Division) Let α and δ be ordinal numbers such that $\delta \neq 0$. Then there are unique ordinal numbers β and γ such that

$$\alpha = \delta\beta + \gamma \quad \text{and} \quad \gamma \in \delta.$$

(In the above theorem, δ is called the divisor, β the quotient and γ the remainder.)

Theorem 1.13.27 (Logarithm) Let α and β be ordinal numbers with $\alpha \neq 0$, and $\beta > 1$. Then there are unique ordinal numbers γ, δ and ρ such that

$$\alpha = \beta^\gamma\delta + \rho \quad \text{and} \quad 0 \neq \delta \in \beta \quad \text{and} \quad \rho \in \beta^\gamma.$$

(In the above theorem, β is called base, γ the logarithm, δ the coefficient, and ρ the remainder.)

EXERCISE

1. Let A and B be equinumerous, and let $(A, \leq)$ and $(B, \leq')$ be well-ordered. Prove that $(A, \leq)$ and $(B, \leq')$ are isomorphic.

2. Let the set $(\mathbb{N}_0, \leq)$ be well-ordered as $0 \leq 1 \leq 2 \leq 3 \leq \ldots$, and as $(\mathbb{N}_0, \leq')$, where

   $$1 \leq' 3 \leq' \ldots \leq' 0 \leq' 2 \leq' 4 \leq' \ldots$$

   Prove that $(\mathbb{N}_0, \leq)$ and $(\mathbb{N}_0, \leq')$ are not isomorphic. Also, show that $\text{ord}(\mathbb{N}_0, \leq) < \text{ord}(\mathbb{N}_0, \leq')$.

3. Let α, β, γ be any ordinal numbers. Prove the following statements.
   (a) $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$. (Associative law)
   (b) $\beta < \gamma$ if and only if $\alpha + \beta < \alpha + \gamma$.
   (c) $\alpha + \beta = \alpha + \gamma$ implies that $\beta = \gamma$. (Left cancellation law)
   (d) Show that $\beta + \alpha = \gamma + \alpha$ does not imply that $\beta = \gamma$.
   (e) $\alpha + 0 = \alpha = 0 + \alpha$.
   (f) $\alpha < \beta$ if and only if $\alpha + 1 \leq \beta$.
   (g) $\alpha + 1 > \alpha$.

4. Let α, β, γ be any ordinal numbers. Prove the following statements.
   (a) $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. (Associative law)
   (b) $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$. (Left distributive law)
   (c) Show, by an example that right distributive law does not hold, that is,
   $$(\alpha + \beta)\gamma \neq \alpha\gamma + \beta\gamma$$
   (d) If $\alpha < \beta$, and $\gamma > 0$, then $\gamma\alpha < \gamma\beta$.
   (e) $\gamma\alpha = \gamma\beta$ implies that $\alpha = \beta$. (Left cancellation law)
   (f) Show that the right cancellation does not hold, that is,
   $$\alpha\gamma = \beta\gamma \text{ does not necessarily imply that } \alpha = \beta.$$
   (g) (i) $0\alpha = 0 = \alpha 0$.
   (ii) $1\alpha = \alpha = \alpha 1$.
   (iii) $\alpha\beta = 0$ if and only if $\alpha = 0$, or $\beta = 0$.

5. Verify that
   (a) $\omega 2 = \omega + \omega$.
   (b) $n\omega = \omega$ for any finite ordinal $n \neq 0$.

6. Show that $0+1+2+\cdots = \omega$. Use the following definition: Let $\{\lambda_i\}_{i\in I}$ be a well-ordered set of ordinal numbers, and $\lambda_i = \text{ord}(A_i)$. Then $\sum_{i\in I} \lambda_i = \text{ord}\left(\bigcup_{i\in I}(A_i \times \{i\}\right)$.

7. Prove the following:
   (a) If $\alpha < \beta$, then $\alpha^+ < \beta^+$ for all ordinals α and β.
   (b) If $\beta > 0$, then $\alpha < \alpha + \beta$ for all ordinals α and β.
   (c) ω^n is a countable set for each $n \in \omega$.

8. Let α, β and γ be any ordinal numbers. The following statements hold.
   (a) If $\beta > 0$, then $0^\beta = 0$, $1^\beta = 1$, $\alpha^1 = \alpha$.
   (b) If $\beta > 0$, then $\alpha^\beta \geq \alpha$.

(c) If $\alpha > 0$, and λ is a limit ordinal, then $\alpha^{\lambda} = \bigcup\{\alpha^{\beta} : \beta < \lambda\}$.

(d) $\alpha^{\beta} = 0$ if and only if $\alpha = 0$, and $\beta > 0$.

(e) $\alpha^{\beta} = 1$ if and only if $\beta = 0$, or $\alpha = 1$.

(f) If $\alpha > 1$ and $\beta < \gamma$, then $\alpha^{\beta} < \alpha^{\gamma}$.

(g) If $\alpha > 1$, and $\alpha^{\beta} = \alpha^{\gamma}$, then $\beta = \gamma$.

(h) If $\alpha \geq 1$, and $\alpha^{\beta} < \alpha^{\gamma}$, then $\beta < \gamma$.

(i) If $\alpha > 1$, then $\alpha^{\beta} \geq \beta$.

(j) $\alpha^{\beta+\gamma} = \alpha^{\beta}\alpha^{\gamma}$.

(k) $(\alpha^{\beta})^{\gamma} = \alpha^{\beta\gamma}$.

1.14 Vector Spaces and Normed Spaces

An element $(x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ can also be regarded as a vector, interpreted as a directed line segment joining the origin $(0, 0, \ldots, 0)$ in $\mathbb{R}^n$ with the point $(x_1, x_2, \ldots, x_n)$, and thus $\mathbb{R}^n$ is thought of as a set of vectors. There is a natural way of adding any two vectors in $\mathbb{R}^n$ and multiplying a vector in $\mathbb{R}^n$ by a number a (called scalar) in the field $\mathbb{R}$:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$
$$a(x_1, x_2, \ldots, x_n) = (ax_1, ax_2, \ldots, ax_n) \quad \text{for all } a \text{ in } \mathbb{R}.$$

These operations called, respectively, vector addition and scalar multiplication satisfy a number of properties which are abstracted to arrive at the following definition.

Definition 1.14.1 Let V be a non-empty set, and F be a field. We define two binary operations $+ : V \times V \to V$ called vector addition and $\cdot : F \times V \to V$ called scalar multiplication, and write $+(x, y)$ as $x + y$ and $\cdot(a, x)$ as ax. Then V is said to be a vector space over the field F, denoted by $V(F)$, if

(i) $x + (y + z) = (x + y) + z$ for all x, y and z in V,

(ii) there exists an element 0 in V such that $x + 0 = 0 + x = x$ for all x in V (0 is called a zero element in V),

(iii) for each element x in V there exists an element, denoted by $-x$, in V such that $x + (-x) = (-x) + x = 0$. ($-x$ is called the additive inverse of x)

(iv) $x + y = y + x$ for all x and y in V,

(v) $a(x + y) = ax + bx$ for all a in F, and x and y in V,

(vi) $(a + b)x = ax + bx$ for all a and b in F, and x in V,

(vii) $(ab)x = a(bx)$ for all a and b in F, and x in V,

(viii) $1x = x$ for all x in V, where 1 is the unit element in F.

The elements of V are usually called vectors, and the elements of F, scalars.

The set $\mathbb{R}^n$ is a vector space over the field of real numbers $\mathbb{R}$. $\mathbb{C}^n$ is a vector space over the field $\mathbb{R}$ (or $\mathbb{C}$) if we interpret $x+y$ and ax as ordinary addition and multiplication (a is real or complex), and define addition and scalar multiplication in $\mathbb{C}^n$ as in $\mathbb{R}^n$:

$$(x_1,\ldots,x_n)+(y_1,\ldots,y_n)=(x_1+y_1,\ldots,x_n+y_n),$$
$$a(x_1,\ldots,x_n)=(ax_1,\ldots,ax_n).$$

There are some elementary properties that each vector space satisfies:

1. There is a unique zero element in V.
2. There is a unique additive inverse of each element in V.
3. $x+y=z+y$ implies that $x=z$.
4. $a0=0$ (two occurrences of zero element in this equality denote the zero element in V),
5. $0x=0$ (0 in the left of this equality is the additive identity in F, and 0 on the right is the zero element in V).
6. $(-1)x=-x$.

It is often convenient to write $x-y$ for $x+(-y)$.

Let V be a vector space over a field F. A non-empty subset M of V is called a subspace of V if with respect to the operations in V, M itself is a vector space over F. It is easily seen that M is a subspace of V if and only if $x+y\in M$ for all x and y in M, and $ax\in M$ for all $a\in F$, and for $x\in M$.

The zero element of V is the zero element of M since $0x=0$. If x is an element in M, then its additive inverse when x is considered as an element of V, and when x is considered as an element of M coincide.

A linear combination of vectors $x_1,x_2,\ldots,x_n$ in a vector space V over a field F is an expression of the form

$$a_1x_1+a_2x_2+\cdots+a_nx_n$$

where $a_1,a_2,\ldots,a_n$ are scalars from F. A linear span of a subset M of V, written span (M), is the set of all linear combination of the vectors in M, that is,

$$\text{span}(M)=\{a_1x_1+\cdots+a_nx_n : x_1,\ldots,x_n\in M,\quad n\in\mathbb{Z}_+,\text{ and } a_1,\ldots,a_n\in F\}.$$

It is readily seen that span (M) is a subspace of V; Infact, it coincides with the intersection of all the subspaces of V that contain M. We say that span (M) is the subspace of V generated by M.

The n vectors $x_1,x_2,\ldots,x_n$ in $V(F)$ are said to be linearly independent if for every set of scalars $a_1,a_2,\ldots,a_n$ in F

$$a_1x_1+\cdots+a_nx_n=0 \text{ implies that } a_1=a_2=\cdots=a_n=0$$

otherwise the vectors $x_1,x_2,\ldots,x_n$ are said to be linearly dependent, that is, the vectors $x_1,\ldots,x_n$ are linearly dependent if there exists a set of scalars $a_1,\ldots,a_n$ not all zero such that $a_1x_1+\cdots+a_nx_n=0$.

In general, a subset S of $V(F)$ is said to be linearly independent if for each $n \in \mathbb{Z}_+$, the vectors in every subset of S having n elements are linearly independent, otherwise S is said to be linearly dependent.

It may be noted that any subset of a linearly independent set is linearly independent, and super set of a linear dependent set is linearly dependent.

A subset S of $V(F)$ is said to be a basis of the vector space V if S is linearly independent, and span $(S) = V$. If S is a finite basis having n elements of a vector space $V(F)$, then V is said to be finite dimensional vector space, and the dimension of V, written $\dim V$, is defined to be n, that is, $\dim V = n$, otherwise V is said to be infinite dimensional vector space.

For example, $\mathbb{R}^n$ is a finite dimensional vector space over the field of real numbers $\mathbb{R}$ since the subset $S = \{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$ of $\mathbb{R}^n$ having n n-tuples is a basis of $\mathbb{R}^n$; the ith tuple in S consists of 1 in the ith co-ordinate and zero elsewhere. Thus, the $\dim(\mathbb{R}^n) = n$.

There are many infinite dimensional vector spaces. Let P be the set of all polynomials with real coefficients defined on $[0, 1]$ and $\mathcal{C}[0, 1]$ be the set of all continuous real valued functions defined on $[0, 1]$. Let $\mathbb{R}$ be the field of real numbers. We define the vector addition and scalar multiplication in both P and $\mathcal{C}[0, 1]$ pointwise as follows: $(f+g)(x) = f(x)+g(x)$ and $(af)(x) = af(x)$ for all $a \in \mathbb{R}$. With respect to these operations both P and $\mathcal{C}[a, b]$ are vector spaces over $\mathbb{R}$. The set of vectors $S = \{1, x, x^2, \ldots\}$ is linearly independent, and

$$\text{span}(S) = P \subsetneqq \mathcal{C}[0, 1].$$

Thus, both P and $\mathcal{C}[0, 1]$ are infinite dimensional vector spaces over the field $\mathbb{R}$.

Note that a vector space always contains the zero element, and a set V that contains only the zero element is a vector space over any field F. If $V = \{0\}$, then by definition V is finite dimensional, and $\dim V$ is defined to be zero, that is, $\dim V = 0$.

It is readily seen that a subspace of a finite dimensional vector space is finite dimensional.

Next we define a normed vector space. The field F in the following is either a real field $\mathbb{R}$ or the field of complex number $\mathbb{C}$.

Definition 1.14.2 Let V be a vector space over a field F. Let $||\ ||: V \to R$ be a mapping (We write $||\ ||(x) = ||x||$.) satisfying the following properties:

(i) $||x|| \geq 0$,

(ii) $||x|| = 0$ if and only if $x = 0$,

(iii) $||ax|| = |a|||x||$,

(iv) $||x + y|| \leq ||x|| + ||y||$ (triangle inequality)

for any x, y in V, and $a \in F$. Then the mapping $||\ ||$ is called a norm on V, and $(V, ||\ ||)$ is called a normed vector space or a normed linear space. We read $||x||$ as "norm of x".

The vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are normed vector spaces if we define the norm by

$$||x|| = \left[\sum_{i=1}^{n} |x_i|^2\right]^{\frac{1}{2}} \quad \text{for all } x = (x_1, \ldots, x_n) \in \mathbb{R}^n \text{ or } \mathbb{C}^n.$$

The vector space $\mathcal{C}[0, 1]$ defined earlier in this section is a normed vector space with the norm defined by

$$||f|| = \sup\{|f(x)| : x \in [0, 1]\}.$$

The following definition of a convex set is often used.

A subset A of a vector space X is said to be a convex set in X, if for all x, y in A the set

$$[x, y] = \{z \in X : z = tx + (1 - t)y \quad \text{where } 0 \leq t \leq 1\}$$

is contained in A. The set $[x, y]$ is called a (closed) straight line segment; x and y are called end points, and any other point in $[x, y]$ is called an interior point of $[x, y]$.

EXERCISE

1. Let F be a real or complex field, and P be the set of all polynomials over F. With usual addition of polynomials, and multiplication of a polynomial by a scalar (an element of F), show that P is a vector space over F.

2. Let P_n be the set of all polynomials of degree less than n (n is a positive integer) over a field of real or complex numbers. Using the usual operations of addition of two polynomials and scalar multiplication, show that P_n is a vector space over F. The zero element is the polynomial identically zero: $p(x) = 0$ for all x; in the usual discussion of degree, the degree of this polynomial is not defined!

3. In a vector space V over a field F, show the following:

   (a) $a.0 = 0$ for all $a \in F$.

   (b) $0x = 0$ for all $x \in V$.

   (c) $(-a)x = -(ax)$ for all $x \in V$, and $a \in F$.

   (d) If $x \neq 0$, then $ax = 0$ implies that $a = 0$.

4. Let $\mathbb{R}$ be the field of real numbers. Prove that the vectors $(1, 1, 0, 0)$, $(0, 1, -1, 0)$ and $(0, 0, 0, 3)$ in $\mathbb{R}^4$ are linearly independent.

5. Extend the set $S = \{(1, 1, 0)\}$ to form a basis of the vector space $\mathbb{R}^3$ over $\mathbb{R}$.

6. Show that a field F is a vector space over F. If F is a field, then F^n is a vector space over F with addition and scalar multiplication defined similarly as in $\mathbb{R}^n$:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$
$$a(x_1, x_2, \ldots, x_n) = (ax_1, ax_2, \ldots, ax_n).$$

7. Show that the set $\mathbb{R}$ of real numbers with usual addition and multiplication of a real number by a rational number is a vector space over the field of rational number $\mathbb{Q}$.

8. Is it true that x, y and z are linearly independent vectors, then so are $x + y$, $y + z$ and $z + x$?

9. Is there a set of three linearly independent vectors in $\mathbb{C}^3$?

10. Is the set of polynomials $1, t, t^2$ and $1 + t + t^2$ in P of Exercise 1 linearly dependent? Prove that the set of any of these three is linearly independent.

11. If x, y and z are vectors such that $x+y+z=0$, then x and y span the same subspace as y and z.

12. Prove that every vector space has a basis (This requires the axiom of choice see [22]).

13. What is the dimension of the vector space $\mathbb{C}$ of all complex number considered over the field of real numbers?

14. Is the set $\mathbb{R}$ of real numbers a finite dimensional vector space over the field of rational numbers?

15. Show that an arbitrary intersection of subspaces of a vector space is a subspace.

16. (i) Show that $\mathbb{R}^n$ is a vector space over the field of real numbers $\mathbb{R}$ with addition and scalar multiplication defined by

$$x+y=(x_1+y_1, x_2+y_2, \ldots, x_n+y_n),$$
$$ax=(ax_1, ax_2, \ldots, ax_n)$$

for all $x=(x_1, x_2, \ldots, x_n)$, $y=(y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, and $a \in \mathbb{R}$. Also show that for each real number p such that $1 \leq p < \infty$, the mapping

$$||x||_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{\frac{1}{p}}$$

is a norm on $\mathbb{R}^n$ so that $\mathbb{R}^n$ is a normed vector space over $\mathbb{R}$.

(ii) In part (i), replace $\mathbb{R}^n$ by $\mathbb{C}^n$, show that $\mathbb{C}^n$ is a normed vector space over $\mathbb{R}$.

(iii) In part (i); replace $\mathbb{R}^n$ by $\mathbb{C}^n$ and the field $\mathbb{R}$ by the field of complex numbers $\mathbb{C}$. Show that $\mathbb{C}^n$ is a normed vector space over $\mathbb{C}$.

17. Let $\mathbb{R}^n$ and $\mathbb{C}^n$ be vector spaces in Exercise 16. Show that the mapping $||x||_\infty = \max\{|x_i|\}$ is a norm.

18. Let p be a real number such that $1 \leq p < \infty$. Let l_p = set of all sequences in $\mathbb{R}$ (or $\mathbb{C}$) such that $\sum_{i=1}^{\infty} |x_i|^p$ converges. Show that l_p is a vector space over the field $\mathbb{R}$ (or $\mathbb{C}$) if addition and scalar multiplication are defined by

$$x+y=\langle x_i+y_i\rangle, \quad ax=\langle ax_i\rangle$$

for all $x=\langle x_i\rangle$, $y=\langle y_i\rangle$ in l_p and $a \in \mathbb{R}$ (or $\mathbb{C}$). (Note that when l_p is the set of sequences in $\mathbb{R}$, then the field to be taken is $\mathbb{R}$.

Also show that the mapping $||x||_p = [\sum |x_i|^p]^{\frac{1}{p}}$ is a norm on l_p.

19. Let l_∞ be the set of bounded sequences in $\mathbb{R}$ or $\mathbb{C}$. Then l_∞ is a vector space with respect to addition and scalar multiplication defined by

$$x+y=\langle x_i+y_i\rangle, \quad ax=\langle ax_i\rangle$$

for all $x=\langle x_i\rangle$ and $y=\langle y_i\rangle$ in l_∞ and $a \in \mathbb{R}$ or $\mathbb{C}$. Show that

$$||x||_\infty = \sup\{|x_i| : i \in \mathbb{Z}_+\}$$

is a norm on l_∞.

20. Show that the set P of all polynomials with real coefficients defined on $[a, b]$ with usual addition and scalar multiplication is vector space over the field of real numbers $\mathbb{R}$, and

$$||p|| = \sup\{|p(x)| : x \in [a, b]\}$$

is a norm on P.

1.15 Real Sequences

Recall that a real sequence s is a function from $\mathbb{N}$ (or $\mathbb{N}_0$) to the set of real numbers $\mathbb{R}$, and if $S(n) = x_n$, then the sequence s is usually denoted in terms of its values as $\langle x_n\rangle_{n=1}^{\infty}$ or simply as $\langle x_n\rangle$, where it is understood that n takes the values $1, 2, \ldots$ in $\mathbb{N}$ (or $\mathbb{N}_0$). For example, if $s : \mathbb{N} \to \mathbb{R}$ is defined by $s(n) = \frac{1}{n}$, then the sequence s is written as $\langle \frac{1}{n}\rangle$. Now, let $\phi : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function then the composition $s \circ \phi$ is a sequence which is called a subsequence of s. Thus, if $S(n) = x_n$ and $\phi(k) = n_k$, then $(s \circ \phi)(k) = s(\phi(k)) = s(n_k) = x_{n_k}$, and $s \circ \phi$ is the subsequence $\langle x_{n_k}\rangle_{k=1}^{\infty}$. Since ϕ is strictly increasing, it follows that $k \leq n_k$ for all k.

A real sequence $\langle x_n\rangle$ is said to be bounded if the set of points of the sequence $\langle x_n\rangle$ is a bounded set, or equivalently, for some real number k, $|x_n| \leq k$, for all n. The sequence $\langle x_n\rangle$ is said to be bounded from above (below) if its set of points has an upper (respectively, lower) bound. It is readily seen that a real sequence is bounded if and only if it is bounded from above and from below.

A concept that is of fundamental importance in mathematical analysis is that of convergence of a sequence.

Definition 1.15.1 A real sequence $\langle x_n\rangle$ converges to a real number x if for each real number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that $|x_n - x| < \epsilon$ for all $n \geq m$.

When a sequence $\langle x_n\rangle$ converges to x, then x is called a limit of $\langle x_n\rangle$, and we write this fact as $\lim_{n\to\infty} x_n = x$, or as $\langle x_n\rangle \to x$. It can be easily shown that a sequence cannot converge to more than one point, and we use the phrase "the limit" of a sequence. For example, the sequence $\langle \frac{1}{n}\rangle$ converges to 0 in $\mathbb{R}$, and so $\lim_{n\to\infty} \frac{1}{n} = 0$.

A real sequence $\langle x_n\rangle$ is said to be a Cauchy sequence if for each real number $\epsilon > 0$, there exists $p \in \mathbb{Z}_+$ such that $|x_n - x_m| < \epsilon$ for all n, $m \geq p$.

In any elementary course of real analysis, the following equivalence is established: A real sequence $\langle x_n\rangle$ is convergent if and only if it is a Cauchy sequence.

Sometimes, in the extended real number system we use the following definitions.

A real sequence $\langle x_n\rangle$ is said to diverge to $+\infty$ ($-\infty$) if for each real number k, there exists an $m \in \mathbb{Z}_+$ such that

$$x_n \geq k \text{ (respectively,} x_n \leq k) \text{ for all } n \geq m.$$

Thus, the sequence $\langle -n\rangle$ diverges to $-\infty$, and the sequence $\langle n\rangle$ diverges to $+\infty$. A sequence which is not bounded from above need not diverge to $+\infty$. For example, the sequence $\langle 1, 2, 1, 3, 1, 4, 1, \ldots\rangle$ does not diverge to $+\infty$.

A real sequence s is said to be monotonically increasing if s is a monotonically increasing function that is $m \leq n$ implies that $s(m) \leq s(n)$. Similarly, if $m \leq n$ implies that $s(m) \geq s(n)$, then s is called monotonically decreasing.

A monotonically increasing and bounded from above sequence is easily seen to be a convergent sequence.

There are two concepts known as "limit superior" and "limit inferior" of a sequence which are related to the concept of convergent sequences. Roughly, the limit superior of a sequence $\langle x_n \rangle$ is a measure of "how large x_n can be when n is large". More formally, we have the following.

Let $\langle x_n \rangle$ be a real sequence which is bounded from above. Let $M_n = \sup\{x_n, x_{n+1}, \ldots\}$. If $\langle M_n \rangle$ converges, then limit superior of $\langle x_n \rangle$, denoted $\lim_{n\to\infty} \sup x_n$, is defined as

$$\lim_{n\to\infty} \sup x_n = \lim_{n\to\infty} M_n$$

and if $\langle M_n \rangle$ diverges to $-\infty$, then $\lim_{n\to\infty} \sup x_n$ is defined to be $-\infty$. If $\langle x_n \rangle$ is not bounded from above, then $\lim_{n\to\infty} \sup x_n$ is defined to be $+\infty$.

In analogous manner, we define the concept of limit inferior of a sequence. Let $\langle x_n \rangle$ be a sequence bounded from below, and $m_n = \inf\{x_n, x_{n+1}, \ldots\}$. If $\langle m_n \rangle$ converges, then limit inferior of $\langle x_n \rangle$, written $\lim_{n\to\infty} \inf x_n$, is defined to be the limit $\lim_{n\to\infty} m_n$. If $\langle m_n \rangle$ diverges to $+\infty$, then $\lim_{n\to\infty} \inf x_n$ is defined to be $+\infty$, and if $\langle x_n \rangle$ is not bounded from below, then $\lim_{n\to\infty} \inf x_n$ is defined to be $-\infty$. Note that the sequence $\langle M_n \rangle$ is monotonically decreasing, thus it is either convergent or it diverges to $-\infty$, and $\langle m_n \rangle$ is monotonically increasing, and so it is either convergent or diverges to $+\infty$. For example, for the sequence $\langle 1, 2, 1, 3, 1, 4, 1, \ldots \rangle$ the limit inferior is 1, and the limit superior is $+\infty$. Also note that limit inferior and limit superior can always be defined for an arbitrary sequence, and these are unique (extended) real numbers. The following theorem summarises the basic properties of the limit superior and limit inferior of sequences.

Theorem 1.15.2 Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be real sequences. Then the following statements hold.

1. $\lim_{n\to\infty} \inf x_n \leq \lim_{n\to\infty} \sup x_n$.

2. If $\langle x_n \rangle$ is convergent, then $\lim_{n\to\infty} \inf x_n = \lim_{n\to\infty} \sup x_n$.

3. If $\lim_{n\to\infty} \inf x_n = \lim_{n\to\infty} \sup x_n = l$, then $\langle x_n \rangle$ is convergent, and $\lim_{n\to\infty} x_n = l$.

4. If $\langle x_n \rangle$ is bounded, and $\lim_{n\to\infty} \sup x_n = M$, then (i) for any real number $\epsilon > 0$, $x_n < M + \epsilon$ for all but a finite number of values of n; (ii) $x_n > M - \epsilon$ for infinitely many values of n.

5. If $\langle x_n \rangle$ is bounded, and $\lim_{n\to\infty} \inf x_n = m$, then for any real real number $\epsilon > 0$ (i) $x_n > m - \epsilon$ for all but a finite number of values of n; (ii) $x_n < m + \epsilon$ for infinitely many values of n.

6. If $\langle x_n \rangle$ and $\langle y_n \rangle$ are bounded, then

   (a) $\lim_{n\to\infty} \sup(x_n + y_n) \leq \lim_{n\to\infty} \sup x_n + \lim_{n\to\infty} \sup y_n$,

   (b) $\lim_{n\to\infty} \inf x_n + \lim_{n\to\infty} \inf y_n \leq \lim_{n\to\infty} \inf(x_n + y_n)$.

7. If $\lim_{n\to\infty} \inf x_n = \lim_{n\to\infty} \sup x_n = +\infty$, then $\langle x_n \rangle$ diverges to $+\infty$.

EXERCISE

1. Find the limit superior and the limit inferior for the following sequences.

   (a) $\langle \sin(\frac{n\pi}{2}) \rangle_{n=1}^{\infty}$.

   (b) $\langle \left(1 + \frac{1}{n}\right)^n \rangle_{n=1}^{\infty}$.

   (c) For any sequence $\langle x_n \rangle$ of all rational numbers in $]0, 1[$, find the limit superior of $\langle x_n \rangle$ and limit inferior of $\langle x_n \rangle$. Use Theorem 1.15.2 (d) and (e).

2. Show that the sequence $\langle \left(1 + \frac{1}{n}\right)^n \rangle_{n=1}^{\infty}$ is convergent.

3. Let $\langle x_n \rangle$ be a real sequence. Let $S_n = x_1 + x_2 + \cdots + x_n$ for all $n \in \mathbb{Z}_+$. Then the ordered pair $(\langle x_n \rangle, \langle S_n \rangle)$ of sequences $\langle x_n \rangle$ and $\langle S_n \rangle$ is said to be an infinite series usually denoted by $\sum_{n=1}^{\infty} x_n$. The infinite series $\sum_{n=1}^{\infty} x_n$ is said to be convergent if the sequence $\langle S_n \rangle$ is convergent. Show that the infinite series $\sum_{n=1}^{\infty} \frac{1}{n}$ is not convergent.

4. An infinite series $\sum_{n=1}^{\infty} x_n$ is said to be absolutely convergent if the infinite series $\sum_{n=1}^{\infty} |x_n|$ is convergent. Show that $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$ is convergent as well as absolutely convergent.

5. A sequence $\langle f_n \rangle$ of real valued functions defined on a domain $D \subseteq \mathbb{R}$ is said to pointwise convergent on D if the sequence $\langle f_n(x) \rangle$ converges for each $x \in D$. We define a function f for a pointwise convergent sequence $\langle f_n \rangle$ by $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in D$. f is called a pointwise limit of $\langle f_n \rangle$. Show that the sequence $\langle f_n \rangle$ defined by

$$f_n(x) = \begin{cases} 0 & \text{if } x < \frac{1}{(n+1)}, \\ \sin^2\left(\frac{\pi}{x}\right) & \text{if } \frac{1}{n+1} \leq x \leq \frac{1}{n}, \\ 0 & \text{if } \frac{1}{n} < x \end{cases}$$

   is pointwise convergent, and its pointwise limit is a continuous function.

6. For $m = 1, 2, \ldots$ let $f_m(x) = \lim_{n\to\infty} (\cos m!\pi x)^{2n}$. Show that the pointwise limit of the sequence is the everywhere discontinuous function

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational,} \\ 1 & \text{if } x \text{ is rational.} \end{cases}$$

7. A sequence of functions $\langle f_n \rangle$ with domain $D \subseteq \mathbb{R}$ converges uniformly to a function f on D if for every real number $\epsilon > 0$, there is an integer n_0 such that

$$|f_n(x) - f(x)| < \epsilon \text{ for all } n \geq n_0, \text{ and for all } x \in D.$$

   Show that the sequence $\langle f_n \rangle$ in Exercise 5 does not converge uniformly to its pointwise limit.

8. Prove that if $a > 1$, the sequence $\langle x_n \rangle$ defined by

$$x_1 = a, x_2 = a^{x_1}, \ldots x_n = a^{x_{n-1}}, \ldots$$

converges if $a \leq e^{\frac{1}{e}}$.

9. For sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ of real numbers, define relations by $\langle x_n \rangle E \langle y_n \rangle$ if $\lim_{n\to\infty}(x_n - y_n) = 0$, and $\langle x_n \rangle F \langle y_n \rangle$ if $\lim_{n\to\infty}(x_n - y_n) = 1$. Which of E and F is an equivalence relation and/or partial ordering? Give reasons.

Chapter 2

Basic Concepts in Metric Spaces

2.1 Definition and Examples of Metric Spaces

In ancient times, when man learnt to count his sheeps, at the same time, for the fear of losing the sheeps in the wild in dark, he, probably, have thought how far the sheeps might have gone from his hutments. Ever since the concept of distance and its measurement has puzzled the man. However, the effort put in to understand these, have paid rich dividends. The most basic definitions in mathematics were initially understood and described using the concept of a distance. For instance, the concepts such as the limit of a function, the convergence of a sequence and continuity of a function at a point used the notion of distance for their description. The definition that follows captures the most fundamental features of the concept of distance as we know it in our daily life. These are:

1. The distance between two objects is greater than or equal to zero.

2. The distance between two objects is zero if and only if the two objects are same.

3. The distance of any object x from another object y is same as the distance of y from x.

4. If there are three objects x, y and z, then the distance between x and y is less than or equal to the sum of the distance between x and z and the distance between y and z.

Definition 2.1.1 Let X be a non-empty set, and $d : X \times X \to \mathbb{R}$ be a mapping having the following properties.

1. $d(x, y) \geq 0$ for all $x, y \in X$,

2. $d(x, y) = 0$ if and only if $x = y$ for all $x, y \in X$,

3. $d(x, y) = d(y, x)$ for all $x, y \in X$,

4. $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y and $z \in X$ (Triangle inequality).

Such a mapping d is called a metric on X, and the pair (X, d) is called a metric space.

The mapping d in the above definition is sometimes called the distance function or simply the distance. Also, when d is understood, we say that X is a metric space. The set $\mathbb{R}$ in the above definition is assumed to be the ordered complete field of real numbers.

In most of the examples given below, it is straightforward to check the first three conditions of a metric, and we leave them for the reader.

Example 2.1.2 Let $X = \mathbb{R}$ and $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a mapping defined by

$$d(x, y) = |x - y| \quad \text{for all } x, y \in \mathbb{R}.$$

Then d satisfies the conditions for being a metric on $\mathbb{R}$.

1. $d(x, y) = |x - y| \geq 0$ for all $x, y \in \mathbb{R}$.
2. $d(x, y) = 0$ if and only if $|x - y| = 0$ if and only if $x = y$ for all $x, y \in \mathbb{R}$.
3. $d(x, y) = |x - y| = |y - x| = d(y, x)$ for all $x, y \in \mathbb{R}$.
4. Let x, y and z be any three elements in $\mathbb{R}$. Then

$$\begin{aligned} d(x, y) &= |x - y| \\ &= |x - z + z - y| \\ &\leq |x - z| + |z - y| \\ &= d(x, z) + d(z, y). \end{aligned}$$

Therefore, d is a metric on $\mathbb{R}$, and $(\mathbb{R}, d)$ is a metric space. This metric space is called the real line, and it is the source of many examples. When d is understood, $(\mathbb{R}, d)$ is denoted by $\mathbb{R}$.

Example 2.1.3 Let $\mathbb{C}$ be the field of complex numbers. Let us define a mapping $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ by

$$d(z_1, z_2) = |z_1 - z_2| \quad \text{for all } z_1, z_2 \in \mathbb{C}.$$

It is easily verified that d is a metric on $\mathbb{C}$. The metric space $(\mathbb{C}, d)$ is called the complex plane. The complex plane is usually written as $\mathbb{C}$.

Example 2.1.4 Let X be a non-empty set, and $d : X \times X \to \mathbb{R}$ be a mapping defined by

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Let x, y and z be any three elements in X. If $x = y$, then

$$d(x, y) = 0 \leq d(x, z) + d(z, y).$$

If $x \neq y$, then z is either different from x or from y. Thus, either $d(x, z) = 1$ or $d(z, y) = 1$. In either case, we have

$$d(x, y) \leq d(x, z) + d(z, y).$$

This proves triangle inequality. Thus, d is a metric on X, known as discrete metric on X. The metric space (X, d) is known as discrete metric space.

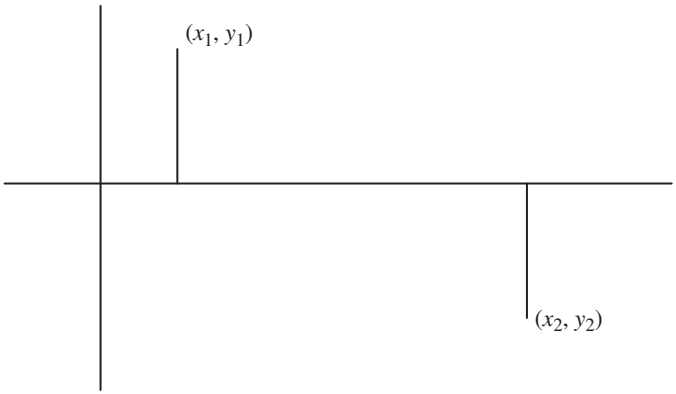


Fig. 2.1

Example 2.1.5 Let $X = \mathbb{R} \times \mathbb{R}$, and define $d : X \times X \to \mathbb{R}$ by

$$d(z_1, z_2) = \begin{cases} |y_1 - y_2| & \text{if } x_1 = x_2, \\ |y_1| + |y_2| + |x_1 - x_2| & \text{if } x_1 \neq x_2, \end{cases}$$

for all $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ in X. This metric is known as "river metric" : Suppose that there is a wooded country in which a river flows along the line $y = 0$. A habitant at (x_1, y_1) can reach another habitant at (x_2, y_2) by walking first along the line $x = x_1$ to the river, then swim along the river to get to the line $x = x_2$, and finally move along the line $x = x_2$ to reach the point (x_2, y_2) (Fig. 2.1). The total distance travelled

$$= |y_1| + |x_1 - x_2| + |y_2|.$$

In the next example, the following inequality, known as the Minkowski's inequality is needed which, in turn needs the following notion of a convex function.

Let$[a, b]$ be an interval in the real line $\mathbb{R}$. A function $f : [a, b] \to \mathbb{R}$ is convex if for any two points x_1 and x_2 in $[a, b]$,

$$f(tx_2 + (1 - t)x_1) \leq tf(x_2) + (1 - t)f(x_1)$$

whenever $0 \leq t \leq 1$.

It is an elementary theorem in real analysis that a function $f : [a, b] \to \mathbb{R}$ is convex on $[a, b]$ if $f''(x) > 0$ for all $x \in [a, b]$. Here $f''(x)$ denotes the second order derivative of f at x.

Lemma 2.1.6 (Minkowski's inequality) Let $a_1, \ldots, a_n$; $b_1, \ldots, b_n$ be real (or complex) numbers, and p a real number such that $p \geq 1$. Then

$$\left\{\sum_{i=1}^{n} |a_i + b_i|^p\right\}^{1/p} \leq \left\{\sum_{i=1}^{n} |a_i|^p\right\}^{1/p} + \left\{\sum_{i=1}^{n} |b_i|^p\right\}^{1/p}.$$

Proof. For $p \geq 1$, the function $f(x) = x^p$ satisfies the inequality $f''(x) \geq 0$, where $f''(x)$ denotes the second order derivative of f at x, for all x in the set $\{x \in \mathbb{R} : x \geq 0\}$. Thus $f(x)$

is a convex function, that is, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for $x, y \geq 0$. Assuming

$$A = \left\{\sum_{i=1}^{n} |a_i|^p\right\}^{1/p}, \; B = \left\{\sum_{i=1}^{n} |b_i|^p\right\}^{1/p}, \; \lambda = \frac{A}{A+B}, \; x = \frac{|a_i|}{A}, \text{ and } y = \frac{|b_i|}{B}$$

and using the convexity of $f(x)$, we get

$$\left[\frac{|a_i|}{A+B} + \frac{|b_i|}{A+B}\right]^p \leq \left(\frac{A}{A+B}\right)\left(\frac{|a_i|^p}{A^p}\right) + \left(\frac{B}{A+B}\right)\left(\frac{|b_i|^p}{B^p}\right).$$

Therefore,

$$\begin{aligned}\frac{|a_i + b_i|^p}{(A+B)^p} &\leq \left[\frac{|a_i|}{A+B} + \frac{|b_i|}{A+B}\right]^p \\ &\leq \frac{|a_i|^p}{A^{p-1}(A+B)} + \frac{|b_i|^p}{B^{p-1}(A+B)}.\end{aligned}$$

Adding, for $i = 1, 2, \ldots, n$, the above inequalities give

$$\sum_{i=1}^{n} |a_i + b_i|^p \leq \left(\frac{(A+B)^{p-1}}{A^{p-1}}\right)\sum_{i=1}^{n} |a_i|^p + \left(\frac{(A+B)^{p-1}}{B^{p-1}}\right)\sum_{i=1}^{n} |b_i|^p = (A+B)^p.$$

So that

$$\left\{\sum_{i=1}^{n} |a_i + b_i|^p\right\}^{1/p} \leq A + B = \left\{\sum_{i=1}^{n} |a_i|^p\right\}^{1/p} + \left\{\sum_{i=1}^{n} |b_i|^p\right\}^{1/p}.$$

The Minkowski's inequality for infinite sums also holds. The precise statement is as follows:

Lemma 2.1.7 Let $p \geq 1$ be a real number, and $\langle a_n \rangle$ and $\langle b_n \rangle$ be seqeunces of real (or complex) numbers such that the infinite series $\sum_{i=1}^{\infty} |a_i|^p$ and $\sum_{i=1}^{\infty} |b_i|^p$ are convergent. Then the infinite series $\sum_{i=1}^{\infty} |a_i + b_i|^p$ converges, and

$$\left\{\sum_{i=1}^{\infty} |a_i + b_i|^p\right\}^{1/p} \leq \left\{\sum_{i=1}^{\infty} |a_i|^p\right\}^{1/p} + \left\{\sum_{i=1}^{\infty} |b_i|^p\right\}^{1/p}.$$

Proof. For any positive integer n, we have

$$\begin{aligned}\left\{\sum_{i=1}^{n} |a_i + b_i|^p\right\}^{1/p} &\leq \left\{\sum_{i=1}^{n} |a_i|^p\right\}^{1/p} + \left\{\sum_{i=1}^{n} |b_i|^p\right\}^{1/p} \\ &\leq \left\{\sum_{i=1}^{\infty} |a_i|^p\right\}^{1/p} + \left\{\sum_{i=1}^{\infty} |b_i|^p\right\}^{1/p}.\end{aligned}$$

From this inequality, the convergence of $\sum_{i=1}^{\infty} |a_i + b_i|^p$ follows, and also the required inequality is derived.

Example 2.1.8 Let p be a real number such that $1 \leq p < \infty$ and $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$. Let $d_p : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{R}$ be a mapping defined by

$$d_p(x, y) = \left\{ \sum_{i=1}^{n} |x_i - y_i|^p \right\}^{1/p}$$

for all $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{F}$. We prove here only the triangle inequality. Let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, and $z = (z_1, z_2, \ldots, z_n)$ be any three elements in $\mathbb{F}^n$. Let us take

$$a_i = x_i - y_i,\ b_i = y_i - z_i,\ i = 1, 2, \ldots, n$$

in the Minkowski's inequality to get

$$\left\{ \sum_{i=1}^{n} |x_i - z_i|^p \right\}^{1/p} \leq \left\{ \sum_{i=1}^{n} |x_i - y_i|^p \right\}^{1/p} + \left\{ \sum_{i=1}^{n} |y_i - z_i|^p \right\}^{1/p}.$$

That is, $d_p(x, z) \leq d_p(x, y) + d_p(y, z)$. Thus, $(\mathbb{F}^n, d_p)$ is a metric space, denoted by $l_p(n)$. When $p = 2$ and $\mathbb{F} = \mathbb{R}$, the metric space $l_2(n)$, that is, $(\mathbb{R}^n, d_2)$, is called n dimensional Euclidean space, denoted by $\mathbb{R}^n$. When $p = 2$ and $\mathbb{F} = \mathbb{C}$, the metric space $(\mathbb{C}^n, d_2)$ is called n-dimensional unitary space, denoted by $\mathbb{C}^n$.

Example 2.1.9 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and $d_\infty : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{R}$ be a mapping defined by

$$d_\infty(x, y) = \max_i \{|x_i - y_i|\}$$

for all $x = (x_1, x_2, \ldots, x_n)$, and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{F}^n$. Then d_∞ is a metric on $\mathbb{F}^n$. Thus, we get the metric spaces $(\mathbb{R}^n, d_\infty)$ and $(\mathbb{C}^n, d_\infty)$.

In the next three examples, three metrics are defined on the set $\mathcal{C}[a, b]$ of all real valued continuous functions defined on the closed interval $[a, b]$.

Example 2.1.10 We define $d : \mathcal{C}[a, b] \times \mathcal{C}[a, b] \to \mathbb{R}$ by

$$d(f, g) = \sup\{|f(t) - g(t)| : t \in [a, b]\}$$

for all f and g in $\mathcal{C}[a, b]$.

If f and g are two functions, we form their sum and negations pointwise as follows:

$$(f + g)(x) = f(x) + g(x),$$
$$(-f)(x) = -f(x).$$

Now the difference of f and g, denoted by $f - g$, is the sum $f + (-g)$:

$$\begin{aligned}(f - g)(x) &= (f + (-g))(x) \\ &= f(x) + (-g)(x) \\ &= f(x) - g(x).\end{aligned}$$

The origin of $\mathcal{C}[a, b]$ is the function 0 given by

$$0(x) = 0 \quad \text{for all } x \in [a, b].$$

Since a real valued continuous functions defined on $[a, b]$ achieves it supremum at a point in $[a, b]$, the sup (supremum) in the above definition of d can be replaced by max (maximum). As usual, we only establish the triangle inequality.

Let f, g and h be three points in $\mathcal{C}[a, b]$. Then for $t \in [a, b]$,

$$\begin{aligned} |f(t) - g(t)| &\leq |f(t) - h(t)| + |h(t) - g(t)| \\ &\leq \sup_{t\in[a,b]} \{|f(t) - h(t)|\} + \sup_{t\in[a,b]} \{|h(t) - g(t)|\} \\ &= d(f, h) + d(h, g). \end{aligned}$$

So that $d(f, g) \leq d(f, h) + d(h, g)$ and d is a metric on $\mathcal{C}[a, b]$. The metric space $(\mathcal{C}[a, b], d)$ is usually denoted by $\mathcal{C}[a, b]$.

Example 2.1.11 Let $\mathcal{C}[a, b]$ be the set as defined in the preceding example.

Now we define a mapping $d : \mathcal{C}[a, b] \times \mathcal{C}[a, b] \to \mathbb{R}$

$$d(f, g) = \int_a^b |f(x) - g(x)|dx$$

where the integral in the above definition is the Riemann integral. In order to check the triangle inequality, let f, g and h be any three elements in $\mathcal{C}[a, b]$. Then

$$\begin{aligned} d(f, h) &= \int_a^b |f(x) - h(x)|dx \\ &= \int_a^b |f(x) - g(x) + g(x) - h(x)|dx \\ &\leq \int_a^b |f(x) - g(x)|dx + \int_a^b |g(x) - h(x)|dx \\ &= d(f, g) + d(g, h). \end{aligned}$$

Thus, $(\mathcal{C}[a, b], d)$ is a metric space. (This example requires the knowledge of the theory of Riemann integration, and the reader with no prior knowledge of this may skip the example and the next one also.)

Example 2.1.12 A third metric on $\mathcal{C}[a, b]$ is given by

$$d(f, g) = \left[\int_a^b (f(t) - g(t))^2 dt\right]^{1/2}.$$

For proving the triangle inequality, the integral version of the Minkowski's inequality for the case $p = 2$ must first be obtained.

Example 2.1.13 Let X be any non-empty set, and $\mathbb{B}(X)$ be the set of all bounded real valued mappings defined on X. We define $\rho : \mathbb{B}(X) \times \mathbb{B}(X) \to \mathbb{R}$ by the formula

$$\rho(f, g) = \sup\{|f(x) - g(x)| : x \in X\}.$$

Then $(\mathbb{B}(X), \rho)$ is a metric space.

Example 2.1.14 Let $(X_i, d_i),\ i = 1, 2, \ldots, n$ be metric spaces, and $X = \prod_{i=1}^{n} X_i$. We define $d^* : X \times X \to \mathbb{R}$ by

$$d^*(x, y) = \left[\sum_{i=1}^{n} \{d_i(x_i, y_i)\}^2\right]^{1/2}$$

for all $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ in X, and

$$d^{**}(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i)$$

for all $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ in X. Then d^* and d^{**} are metrics on X. The triangle inequality for d^* follows using Minkowski's inequality for $p = 2$:

$$\begin{aligned} d^*(x, z) &\leq \left[\sum_{i=1}^{n} \{d_i(x_i, y_i) + d_i(y_i, z_i)\}^2\right]^{1/2} \\ &\leq \left[\sum_{i=1}^{n} \{d_i(x_i, y_i)\}^2\right]^{1/2} + \left[\sum_{i=1}^{n} \{d_i(y_i, z_i)\}^2\right]^{1/2} \\ &= d^*(x, y) + d^*(y, z). \end{aligned}$$

From a given metric, we can always define a bounded metric (A metric is said to be bounded if it is bounded as a real valued mapping) on the same set as in the following two examples.

Example 2.1.15 Let (X, d) be a metric space, and

$$d_1(x, y) = d(x, y)/(1 + d(x, y)) \text{ for all } x, y \in X.$$

Then d_1 is a metric on X. To see the triangle inequality, let x, y and z be any elements in X. Since d is a metric on X, $d(x, z) \leq d(x, y) + d(y, z)$. Then

$$d(x, z) + d(x, z)d(x, y) + d(x, z)d(y, z)$$

$$\leq d(x, y) + d(y, z) + d(x, z)d(x, y) + d(x, z)d(y, z).$$

So that

$$d(x, z)(1 + d(x, y) + d(y, z)) \leq [d(x, y) + d(y, z)][1 + d(x, z)].$$

That is

$$\begin{aligned} \frac{d(x, z)}{[1 + d(x, z)]} &\leq \frac{(d(x, y) + d(y, z))}{[1 + d(x, y) + d(y, z)]} \\ &\leq d_1(x, y) + d_1(y, z). \end{aligned}$$

Thus, we obtain $d_1(x, z) \leq d_1(x, y) + d_1(y, z)$.

There is alternative proof for the triangle inequality. Let us consider the function $f(u) = \frac{u}{(1+u)}$ for all positive real numbers u. It can be easily seen that f is a non-decreasing function.

Let us take $u = d(x, z)$, and $v = d(x, y) + d(y, z)$. Then $u \leq v$. So that $f(u) \leq f(v)$. That is,

$$\frac{d(x,z)}{1+d(x,z)} \leq \frac{d(x,y)+d(y,z)}{1+d(x,y)+d(y,z)}$$

and the triangle inequality is now immediate.

Example 2.1.16 Let (X, d) be a metric space. Let $\rho(x, y) = \min\{1, d(x, y)\}$. Then ρ is a metric on X.

Example 2.1.17 The following metric known as chordal metric is useful in complex analysis. Let $\mathbb{C}$ be the complex plane, and $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ be defined as

$$d(z, z') = \frac{|z - z'|}{\sqrt{(1 + |z|^2)\,(1 + |z'|^2)}} \text{ for all } z \text{ and } z' \text{ in } \mathbb{C}.$$

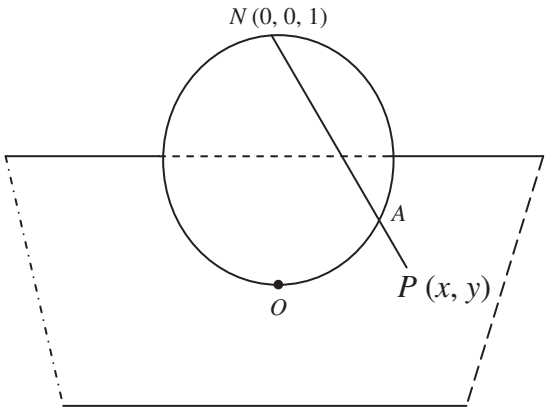


Fig. 2.2

To explain the name chordal of d, we consider a sphere $S : (x - 0)^2 + (y - 0)^2 + (z - 1/2)^2 = 1/4$ of unit diameter placed at the origin O of the complex plane (Fig. 2.2). We identify $\mathbb{C}$ with $\{(x, y, 0) : x, y \in \mathbb{R}\}$. Let $P(x, y)$ be any point in the complex plane representing $z = x + iy$ Now join the north pole $N(0, 0, 1)$ of the sphere to the point P. The line NP intersects S at a point A. In this way, we get a one-to-one correspondence between $\mathbb{C}$ and $S - \{(0, 0, 1)\}$. We investigate this correspondence. The equation of the line NP in $\mathbb{R}^3$ is given by $\{tN + (1 - t)P : -\infty < t < \infty\}$, or by $\{((1 - t)x, (1 - t)y, t) : -\infty < t < \infty\}$.

The value t for which this line intersects the sphere S in a point other than the north pole is given by

$$(1 - t)^2 x^2 + (1 - t)^2 y^2 + (t - 1/2)^2 = \frac{1}{4}$$

or

$$(1 - t)^2 |z|^2 - t(1 - t) = 0.$$

Since $t \neq 1$ (for $t = 1$ the point on the line is N),

$$t = \frac{|z|^2}{|z|^2 + 1}.$$

Thus, the coordinates of the point $A(x_1, x_2, x_3)$ are given by

$$x_1 = \frac{x}{|z|^2 + 1},\ x_2 = \frac{y}{|z|^2 + 1},\ x_3 = \frac{|z|^2}{|z|^2 + 1}.$$

That is,

$$x_1 = \frac{(z + \bar{z})}{2\left(|z|^2 + 1\right)},\ \ x_2 = \frac{(z - \bar{z})}{2\left(|z|^2 + 1\right)},\ \ x_3 = \frac{|z|^2}{|z|^2 + 1}.$$

Given the co-ordinate of the point $A(x_1, x_2, x_3)$, a similar calculation will show that the co-ordinates of the point $P(x, y)$ are given by

$$x = \frac{x_1}{1 - x_3}, \qquad y = \frac{x_2}{1 - x_3}.$$

Now let $A(x_1, x_2, x_3)$ and $A'(x_1', x_2', x_3')$ be the points on S corresponding to the complex numbers $z = x + iy$ and $z' = x' + iy'$ in $\mathbb{C}$, respectively. Then $d(z, z')$ is defined as the length of the chord AA', and hence the name chordal. That is,

$$d(z, z') = \left[(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2\right]^{1/2}.$$

Since A and A' lie on the sphere S, we have

$$[d(z, z')]^2 = x_3 + x_3' - 2(x_1x_1' + x_2x_2' + x_3x_3').$$

Now using the values of x_1, x_2 and x_3, we get

$$d(z, z') = \frac{|z - z'|}{\left[(1 + |z|^2)(1 + |z'|^2)\right]^{1/2}} \text{ for all } z, z' \in \mathbb{C}.$$

Example 2.1.18 We extend the complex plane $\mathbb{C}$ by adding to it the symbol ∞ (infinity), and denote it by $\mathbb{C}_\infty$, that is, $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. Now we extend the chordal metric of the preceding example to $\mathbb{C}_\infty$. First note that as $|z|$ tends to infinity, the point A tends to the north pole. Thus, we identify north pole N in S and ∞ in $\mathbb{C}_\infty$. We define $d_\infty : \mathbb{C}_\infty \times \mathbb{C}_\infty \to \mathbb{R}$ by

$$d_\infty(z, z') = \frac{|z - z'|}{\left[\left(|z|^2 + 1\right)\left(|z'|^2 + 1\right)\right]^{1/2}} \text{ for all } z \text{ and } z' \text{ in } \mathbb{C},$$

$$d_\infty(z, \infty) = \frac{1}{\left[1 + |z|^2\right]^{1/2}} \text{ for all } z \in \mathbb{C},$$

$$d_\infty(\infty, \infty) = 0.$$

Note that $d_\infty(z, \infty)$ is the length of the chord AN, where A is the point on S corresponding to z. That d_∞ is indeed a metric on $\mathbb{C}_\infty$ can now easily be verified. The metric space $(\mathbb{C}_\infty, d_\infty)$ is called extended complex plane. The correspondence between the points of S and $\mathbb{C}_\infty$ is called the stereographic projection.

Example 2.1.19 Let $\mathbb{Q}$ be the set of rational numbers and $d : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ be defined as follows. Let p be a prime number, and for a natural number $n > 0$, define $V_p(n)$ as the exponent of p in the decomposition of n into prime factors. Then for any pair of positive integers m and n, it follows that

$$V_p(mn) = V_p(m) + V_p(n).$$

Now let $x = \pm r/s$ be a non-zero rational number with r and s positive. Then, define $V_p(x) = V_p(r) - V_p(s)$. This definition of $V_p(x)$ does not depend on any particular expression of x as a fraction. For instance $V_2(8/24) = V_2(1/3)$. Also it follows that

$$V_p(xy) = V_p(x) + V_p(y)$$

for all non-zero rational numbers x and y. We define

$$d(x, y) = \begin{cases} p^{-V_p(x-y)} & \text{if } x \neq y, \\ 0 & \text{if } x = y \end{cases}$$

for all x and y in $\mathbb{Q}$. We show that d is a metric on $\mathbb{Q}$. $d(x, y)$ is obviously non-negative. Also note that for $x \neq y$, $p^{-V_p(x-y)}$ is never zero. Thus, $d(x, y) = 0$ if and only if $x = y$. Since $V_p(x - y) = V_p(y - x)$, $d(x, y) = d(y, x)$ follows. The triangle inequality for d will follow if we prove the inequality

$$d(x, z) \leq \max \{d(x, y)\, ,\, d(y, z)\}\, .$$

This inequality obviously holds if any two of x, y and z are equal. So, we assume that x, y and z are different from each other. The above inequality will hold if

$$V_p(x - z) \geq \min \{V_p(x - y),\, V_p(y - z)\}\, .$$

This inequality will hold if for rational numbers $x \neq 0$, $y \neq 0$ and $x - y \neq 0$,

$$V_p(x - y) \geq \min \{V_p(x), V_p(y)\}\, .$$

We can assume that $V_p(x) \geq V_p(y)$. Then the inequality to be proved becomes

$$V_p(x - y) - V_p(y) \geq 0.$$

Now using the equality $V_p(xy) = V_p(x) + V_p(y)$, we have

$$\begin{aligned} V_p(x - y) &= V_p\left(\frac{y(x - y)}{y}\right) \\ &= V_p\left(\frac{x}{y} - 1\right) + V_p(y), \end{aligned}$$

and so the required inequality is

$$V_p\left(\frac{x}{y} - 1\right) \geq 0.$$

Thus, we are required to prove

$$V_p(z - 1) \geq 0$$

for any rational number z such that $z \neq 0,\ z \neq 1$, and $V_p(z) \geq 0$. But then, by the definition of V_p, $z = \pm p^h \cdot r/s$, where $h \geq 0$, and r and s are not divisible by p. Now the denominator in $z-1$ is not divisible by p, and so the inequality $V_p(z-1) \geq 0$ follows from the definition of V_p. Thus, d is a metric on $\mathbb{Q}$, and it is known as p-adic distance. The p-adic distance has several interesting properties.

Example 2.1.20 Let p be a real number such that $1 \leq p < \infty$, and l_p be the set of all real (or complex) sequences $\langle x_n \rangle$ such that $\sum_{i=1}^{\infty} |x_i|^p$ is convergent. Let $d_p : l_p \times l_p \to \mathbb{R}$ be defined by

$$d_p(x,y) = \left[\sum_{i=1}^{\infty} |x_i - y_i|^p\right]^{1/p} \quad \text{for all } x = \langle x_n \rangle \text{ and } y = \langle y_n \rangle \text{ in } l_p.$$

By Lemma 2.1.7, the convergence of the series $\sum_{i=1}^{\infty} |x_i - y_i|^p$ follows. This means that $d_p(x,y)$ is well-defined. To see the triangle inequality for d, let $x = \langle x_n \rangle$, $y = \langle y_n \rangle$, and $z = \langle z_n \rangle$ be any three elements in l_p. Let $a_i = x_i - y_i$, and $b_i = y_i - z_i$ for all $i \in \mathbb{Z}_+$. Substituting these values in Lemma 2.1.7, the required inequality is obtained:

$$\left[\sum_{i=1}^{\infty} |x_i - z_i|^p\right]^{1/p} \leq \left[\sum_{i=1}^{\infty} |x_i - y_i|^p\right]^{1/p} + \left[\sum_{i=1}^{\infty} |y_i - z_i|^p\right]^{1/p}.$$

The metric space (l_p, d) is usually denoted by l_p. The l_2-space is known as Hilbert space.

In the following example, we need the concept of an arbitrary sum of real numbers and its convergence. Let Λ be an arbitrary index set, and $f : \Lambda \to \mathbb{R}$ be a mapping such that $f(\alpha) = x_\alpha$. Then f can be denoted in terms of its values $\{x_\alpha\}_{\alpha \in \Lambda}$. Let Λ_0 is the set of all finite subsets of Λ. Let $S : \Lambda_0 \to \mathbb{R}$ be the mapping defined by

$$S(J) = \sum_{\alpha \in J} x_\alpha \text{ for all } J \in \Lambda_0.$$

Then the ordered pair $\langle f, S \rangle$ is called an infinite series, usually denoted by $\sum_{\alpha \in \Lambda} x_\alpha$. The infinite serves $\langle f, S \rangle$ is said to converge to a real number x if for each real number $\varepsilon > 0$, there exists a set $J_0 \in \Lambda_0$ such that

$$|S(J) - x| < \varepsilon \text{ for all } J \in \Lambda_0 \text{ and } J_0 \subseteq J,$$

that is

$$\left|\sum_{\alpha \in J} x_\alpha - x\right| < \varepsilon \text{ for all } J \in \Lambda_0 \text{ and } J_0 \subseteq J.$$

Example 2.1.21 Let X be a set of cardinality τ, $H(X)$ be the set of all non-negative real valued mappings f on X such that $f(x) \neq 0$ for at most countably many points x in X, and the series $\sum_{x \in X} (f(x))^2$ converges. Let $\rho : H(X) \times H(X) \to \mathbb{R}$ be defined as

$$\rho(f,g) = \left[\sum_{x \in X} (f(x) - g(x))^2\right]^{1/2}.$$

Then ρ is a metric on $H(X)$. The metric space $(H(X), \rho)$ is called generalised Hilbert space of weight τ, and denoted by $H(\tau)$. The construction does not depend in an essential way on the choice of a set X of given cardinality τ.

Example 2.1.22 Let X be the set of real (or complex) sequences, and

$$\sigma(x, y) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{2^i(1 + |x_i - y_i|)}$$

for all $x = \langle x_n \rangle$, and $y = \langle y_n \rangle$ in X. Then (X, d) is a metric space. Since $d(x_i, y_i) = |x_i - y_i|$ is the usual metric on $\mathbb{R}$, by Example 2.1.15, $\rho(x_i, y_i) = d(x_i, y_i)/(1 + d(x_i, y_i))$ is a metric on $\mathbb{R}$, and $\rho(x_i, y_i) \leq 1$. This means that $\sum_{i=1}^{n} \frac{\rho(x_i, y_i)}{2^i} \leq \sum_{i=1}^{n} \frac{1}{2^i}$. Since the series $\sum_{i=1}^{\infty} \frac{1}{2^i}$ is convergent, it follows that $\sigma(x, y)$ is well-defined. Also adding the triangle inequalities

$$\rho(x_i, y_i) \leq \rho(x_i, z_i) + \rho(z_i, y_i)$$

after multiplying by $\frac{1}{2^i}$, we get

$$\sum_{i=1}^{n} \frac{\rho(x_i, y_i)}{2^i} \leq \sum_{i=1}^{n} \frac{\rho(x_i, z_i)}{2^i} + \sum_{i=1}^{n} \frac{\rho(z_i, y_i)}{2^i},$$

where $n \in \mathbb{Z}_+$. Now the triangle inequality for σ follows. This metric space is known as real (complex) Fréchet space if X is the set of real (complex) sequences.

Example 2.1.23 Let X be the set of all bounded real (or complex) sequences. Then for a sequence $\langle x_n \rangle$ in X, the supremum:

$$\sup \{|x_n| \, : \, n \in \mathbb{Z}_+\}$$

exists. So we define

$$d(x, y) = \sup \{|x_n - y_n| : n \in \mathbb{Z}_+\}$$

for all sequences $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in X. $d(x, y)$ is well-defined since

$$\begin{aligned} d(x, y) &= \sup \{|x_n - y_n| : n \in \mathbb{Z}_+\} \\ &\leq \sup \{|x_n| : n \in \mathbb{Z}_+\} + \sup \{|y_n| : n \in \mathbb{Z}_+\} < \infty. \end{aligned}$$

To show that (X, d) is a metric space is left to the reader. This metric space is usually known as l_∞-space.

Example 2.1.24 Let X be the set of all convergent real (or complex) sequences, and $d(x, y) = \sup \{|x_n - y_n| : n \in \mathbb{Z}_+\}$ for all $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in X. Then (X, d) is a metric space.

Example 2.1.25 Let X be an infinite set of cardinality τ, and for each $x \in X$ set $I_x = I \times \{x\}$, where $I =]0, 1]$. Let $P = \bigcup_{x \in X} I_x$, and adjoin a point (0,0) to P. Let $\widetilde{P} = P \cup \{(0, 0)\}$. Let $\rho : \widetilde{P} \times \widetilde{P} \to \mathbb{R}$ be defined as

$$\rho[(t_1, x), (t_2, y)] = \begin{cases} |t_1 - t_2| & \text{if } x = y, \\ |t_1| + |t_2| & \text{if } x \neq y. \end{cases}$$

Then ρ is a metric on $\widetilde{P}$. The metric space $(\widetilde{P}, \rho)$ is denoted by $J(\tau)$, and is known as the hedgehog space of spinness τ.

Example 2.1.26 Let $D = \{z : z \in \mathbb{C} \text{ and } |z| < 1\}$ and let $d : D \times D \to \mathbb{R}$ be defined by

$$d(z, z') = \frac{1}{2} \log \left(\frac{(1 + \phi(z, z'))}{(1 - \phi(z, z'))} \right)$$

where $\phi(z, z') = |(z - z')/(1 - \bar{z}z')|$. Then d is a metric on unit disc D known as the hyperbolic metric. This metric is important in the study of Hyperbolic geometry.

Example 2.1.27 Let τ be any cardinal number, and $D(\tau)$ a set of cardinality τ. Let $\mathbb{B}(\tau)$ denote the set of all sequences $x = \langle x_i \rangle$ with $x_i \in D(\tau)$ for all $i \in \mathbb{N}$. We define $\rho : \mathbb{B}(\tau) \times \mathbb{B}(\tau) \to \mathbb{R}$ by

$$\rho(x, y) = \frac{1}{k} \text{ for } x = \langle x_i \rangle, \text{ and } y = \langle y_i \rangle \text{ in } \mathbb{B}(\tau),$$

where k is the least integer such that $x_k \neq y_k$. Then ρ is a metric on $\mathbb{B}(\tau)$. The metric space $(\mathbb{B}(\tau), \rho)$ is called the Baire space of weight τ. To see the triangle inequality for ρ, let $x = \langle x_i \rangle$, $y = \langle y_i \rangle$, and $z = \langle z_i \rangle$ be any three elements in $\mathbb{B}(\tau)$. Let $\rho(x, z) = 1/k_1$, and $\rho(z, y) = 1/k_2$, where k_1 and k_2 are the least integers as in the definition. Let $k_1 \leq k_2$. Then

$$x_i = z_i \text{ for all } i = 1, 2, ..., k_1 - 1; \; x_{k_1} \neq z_{k_1}, \text{ and}$$
$$z_i = y_i \text{ for all } i = 1, 2, ..., k_2 - 1; \; z_{k_2} \neq y_{k_2}.$$

Therefore, the least integer k for, which $x_k \neq y_k$ is $k = k_1$ and $\rho(x, y) = 1/k_1$. Now it to follows that $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$.

Example 2.1.28 Let X be the set of all real (or complex) sequences, and

$$d(x, y) = \sum_{i=1}^{\infty} \frac{\min \{|x_i - y_i|, 1\}}{i^2}$$

for all $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in X. Then (X, d) is a metric space.

Example 2.1.29 Let X be the set of all real (or complex) sequences, and

$$\rho(x, y) = \sum_{i=1}^{\infty} \frac{A_n |x_i - y_i|}{(1 + |x_i - y_i|)}$$

for all $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in X, and the series $\sum_{n=1}^{\infty} A_n$ is convergent. Then ρ is a metric on X.

Example 2.1.30 Let us consider the set $\mathcal{C}[0, 1]$ of real valued continuous functions defined on $[0, 1]$, and define

$$\rho(f, g) = \int_0^1 |f(t) - g(t)|^2 \, dt.$$

If we now take $f(t) = 0$, $g(t) = t$, and $h(t) = 1$ for all $t \in [0, 1]$, then $\rho(f, g) = 1/3$, $\rho(f, h) = 1$, and $\rho(h, g) = 1/3$. Therefore, the triangle inequality $\rho(f, h) \leq \rho(f, g) + \rho(g, h)$ is violated. Thus, ρ is not a metric on $\mathcal{C}[0, 1]$.

We can generalise Example 2.1.13 as follows:

Example 2.1.31 Let Z be a non-empty set, and (X, d) be a metric space, then the set $\mathbb{B}(Z, X)$ of all bounded mappings of Z into X (see Section 2.4) with the mapping $\rho(f, g) = \sup\{d(f(x), g(x)) : x \in Z\}$ becomes the metric space$\{\mathbb{B}(Z, X), \rho\}$.

Example 2.1.32 Let $(X, \| \|)$ be a normed vector space. Then the mapping $d : X \times X \to \mathbb{R}$ defined by

$$d(x, y) = \|x - y\| \text{ for all } x, y \text{ in } X$$

is easily seen to be a metric on X, and (X, d) is a metric space. The metric d is said to be induced by the norm $\| \|$. It is in this sense a normed vector space can always be considered as a metric space.

EXERCISE

1. Let $X = \mathbb{R}^2$. If p and q are points in $\mathbb{R}^2$, define $d(p, q)$ to be the sum of the usual distance from p to (0,0) and the distance from (0,0) to q whenever p and q do not lie on a line passing through the origin. Otherwise $d(p, q)$ is defined to be the usual distance between p and q. Is d a metric on X?

2. Let X be the set of all differentiable function f whose derivatives f' are continuous on the closed and bounded interval $[a, b]$. Show that the mapping defined by

   $$d(f, g) = |f(a) - g(a)| + \sup\{|f'(x) - g'(x)| : x \in [a, b]\}$$

   is a metric on X.

3. Show that the discrete metric on a vector space $X \neq \{0\}$ can not be obtained from a norm.

2.2 Open Spheres

The concept of an open sphere or an open ball plays a fundamental role in theory of metric spaces.

Definition 2.2.1 Let (X, d) be a metric space, and x_0 be a point in X. Let $\varepsilon > 0$ be a real number. An open sphere with centre at x_0 and radius ε, denoted by $S(x_0, \varepsilon, d)$, is a set defined by

$$S(x_0, \varepsilon, d) = \{x \in X \ : \ d(x, x_0) < \varepsilon\}.$$

An open sphere is sometimes called an open ball. When the metric d is understood from the context, we denote the open sphere $S(x_0, \varepsilon, d)$ simply by $S(x_0, \varepsilon)$.

Definition 2.2.2 Let (X, d) be a metric space, and x_0 be a point in X. Let $\varepsilon > 0$ be a real number. A closed sphere with centre at x_0 and radius ε, denoted by $S[x_0, \varepsilon, d]$, is defined as

$$S[x_0, \varepsilon, d] = \{x \in X : d(x, x_0) \leq \varepsilon\}.$$

Where d is understood, $S[x_0, \varepsilon, d]$ is denoted by $S[x_0, \varepsilon]$.

Examples 2.2.3

1. In the real line $\mathbb{R}$ (Example 2.1.2), the open sphere centered at a point $a \in \mathbb{R}$ and of radius $\varepsilon > 0$ is given by

$$\begin{aligned} S(a, \varepsilon) &= \{x \in \mathbb{R} : d(x, a) < \varepsilon\} \\ &= \{x \in \mathbb{R} : |x - a| < \varepsilon\} =]a - \varepsilon, a + \varepsilon[. \end{aligned}$$

   Thus, the open spheres in the real line $\mathbb{R}$ are open intervals.

2. In two dimensional Euclidean plane $\mathbb{R}^2$ (Example 2.1.8), the open sphere centered at the origin $(0, 0)$ and with radius 1 is given by

$$\begin{aligned} S((0,0), 1) &= \{(x, y) \in \mathbb{R}^2 : d(x, y), (0, 0)) < 1\} \\ &= \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} < 1\}. \end{aligned}$$

   This set is an open disk centered at the origin (dotted region in Fig. 2.3). The open spheres in $\mathbb{R}^2$ are usually called open disks. Thus, the open spheres on $\mathbb{R}^2$ are open disks.

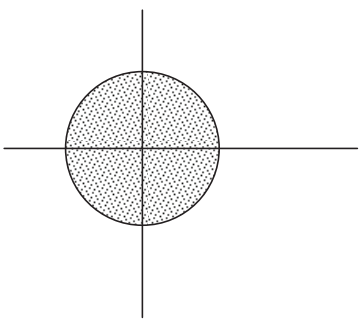


Fig. 2.3

3. In three dimensional Euclidean space $\mathbb{R}^3$ (Example 2.1.8), the open sphere centered at origin $(0, 0, 0)$ and with radius 1 is given by

$$\begin{aligned} S((0,0,0), 1) &= \{(x, y, z) \in \mathbb{R}^3 : d((x, y, z), (0, 0, 0)) < 1\} \\ &= \{(x, y, z) \in \mathbb{R}^3 : \sqrt{x^2 + y^2 + z^2} < 1\} \end{aligned}$$

   This is shown as the dotted region in Fig. 2.4.

   (The word open sphere has been taken from the usual concept of sphere in three dimensional Euclidean space.)

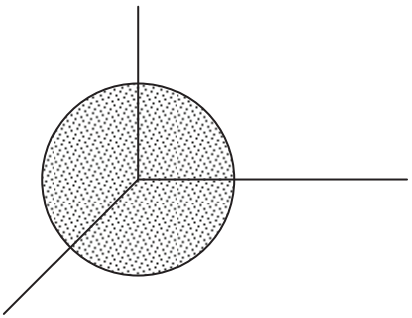


Fig. 2.4

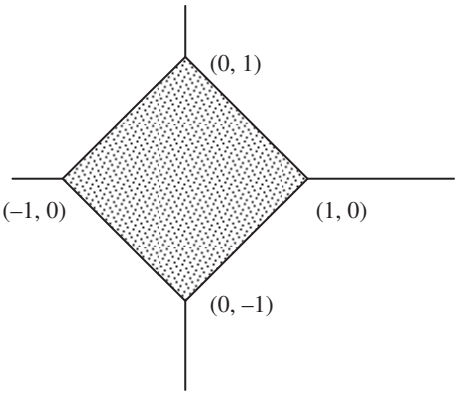


Fig. 2.5

4. Consider the metric space $(\mathbb{R}^2, d_1)$ (Example 2.1.8, $p = 1$). The metric d_1 is given by

$$d_1(x, y) = |x_1 - y_1| + |x_2 - y_2|,$$

where $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R}^2$.

The open sphere $S((0, 0), 1)$ is the set

$$\begin{aligned} S((0,0),1) &= \{(x,y) \in \mathbb{R}^2 : d_1((x,y),(0,0)) < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : |x-0| + |y-0| < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : \pm x \pm y < 1\}. \end{aligned}$$

This is shown as the dotted region in Fig. 2.5.

5. In the metric space $(\mathbb{R}^2, d_\infty)$ (Example 2.1.9), the open sphere $S((0, 0), 1)$ is the set

$$\begin{aligned} S((0,0),1) &= \{(x,y) \in \mathbb{R}^2 : d_\infty((x,y),(0,0)) < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : \max\{|x-0|, |y-0|\} < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : |x-0| < 1 \text{ and } |y-0| < 1\}. \end{aligned}$$

This is shown as the dotted region in Fig. 2.6.

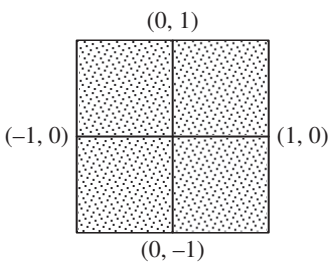


Fig. 2.6

6. In a discrete metric space (X, d), the open sphere centered at $x \in X$ and of radius $\varepsilon (0 < \varepsilon \leq 1)$ is the set

$$S(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\} = \{x\}.$$

In case $\varepsilon > 1$, $S(x, \varepsilon) = X$. Thus, the open spheres in a discrete metric space are either singletons, or coincide with the whole of X.

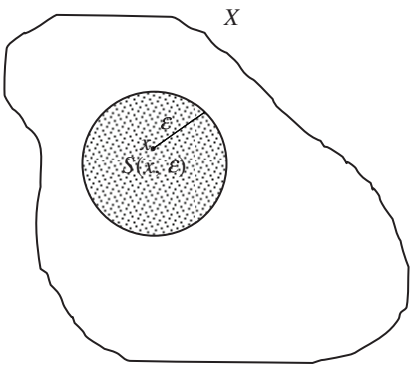


Fig. 2.7

It should be noted that the geometrical depiction of open spheres as above is not a theoretical requirement, and is merely an aid for a beginner to grasp the concept of an open sphere. In any case, such a depiction is not possible in most examples of metric spaces. Fortunately, we would not have many occasions to consider specific sets of open spheres in the development of the theory of metric spaces. However, the reader will find it convenient in general to think an open sphere in a metric space (X, d) as an open disk (Fig. 2.7).

EXERCISE

1. Let $(X, \| \ \|)$ be a normed vector space. Show that the closed unit sphere $S[0, 1]$ in X is a convex set.

2. The sphere $S = \{x \in X : \|x\| = 1\}$ in a normed space X is called the unit sphere. For the norms on $\mathbb{R}^2$ defined by

$$\|x\|_1 = |\xi_1| + |\xi_2|,$$
$$\|x\|_2 = (\xi_1^2 + \xi_2^2)^{1/2},$$
$$\|x\|_4 = (\xi_1^4 + \xi_2^4)^{1/4},$$
$$\|x\|_\infty = \max\{|\xi_1|, |\xi_2|\},$$

for all $x = (\xi_1, \xi_2) \in \mathbb{R}^2$, sketch the unit spheres.

2.3 Neighbourhood of a Point

Definition 2.3.1 Let (X, d) be a metric space, and N be a subset of X. Let x be a point in X. Then N is called a neighbourhood (nbd) of x if there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq N$.

It follows from the above definition that if a point x is not in N, then N can not be a neighbourhood of x, for any open sphere centered at x contains x, and therefore cannot be contained in N.

It may be noted that if (X, d) is a metric space, and x is any point in X, then X is a neighbourhood of x. Thus, X is a neighbourhood of each of its points.

Examples 2.3.2

1. The open interval $]0, 1[$ in the real line $\mathbb{R}$ is a neighbourhood of each of its points. This can be seen as follows. Let x be any point in $]0, 1[$. Then $0 < x < 1$. Let $\varepsilon = \min\{x - 0, 1 - x\}$. Then the open sphere $S(x, \varepsilon) = [x - \varepsilon, x + \varepsilon[$ is contained in $]0, 1[$. Since x is arbitrary point in $]0, 1[$, it follows that $]0, 1[$ is a neighbourhood of each of its points.

2. Again let us consider the set $\mathbb{Q}$ of rational number in the real line $\mathbb{R}$. The set $\mathbb{Q}$ is not a neighbourhood of any point in $\mathbb{R}$. Any open sphere centered at $x \in \mathbb{R}$ is an open interval in $\mathbb{R}$, and contains irrational numbers. So there is no open sphere centered at x and contained in $\mathbb{Q}$.

3. The closed interval $[0, 1]$ in the real line $\mathbb{R}$ is not a neighbourhood of 0. Any open sphere centered at 0 is of the form $]-\varepsilon, \varepsilon[$ which is not contained in $[0, 1]$. Similarly, $[0, 1]$ is not a neighbourhood of 1.

4. Consider the set $N = \{(x, 3x) : x \text{ is a real number}\}$ in two dimensional Euclidean space $\mathbb{R}^2$. In fact, N is the line $y = 3x$. It can be seen that N is not a neighbourhood of any point in $\mathbb{R}^2$.

5. The set $N = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$ is not a neighbourhood of any point of the form $(x, 0)$ and $(0, x)$, where x is a real number. Any open sphere $S((0, x), \varepsilon)$ which is an open disk, conains points of $\mathbb{R}^2$ which are not in N. If (x, y), $x \neq 0$, $y \neq 0$ is any point in N, then N is a neighbourhood of (x, y).

6. The set $\mathbb{Q}\times\mathbb{Q}$ in $\mathbb{R}^2$ is not a neighbourhood of any point in $\mathbb{R}^2$. $\mathbb{Q}\times\mathbb{Q} = \{(x, y) \in \mathbb{R}^2 : x$ and y are rational numbers$\}$.

7. In a discrete metric space (X, d), every subset of X is a neighbourhood of each of its points.

8. In any metric space, every open sphere is a neighbourhood of each of its points.

9. Consider the metric space $\mathcal{C}[0, 1]$ (Example 2.1.10). Let $f \in \mathcal{C}[0, 1]$ be such that $f(x) = 0$ for all x in $[0, 1]$. Let $N = \{g \in \mathcal{C}[0, 1] : -1 < g(x) < 1$ for all x in $[0, 1]\}$. Then N is a neighbourhood of f. Let h be a function which takes the value 1 at some point in $[0, 1]$. It is easily seen that N is not a neighbourhood of h.

Theorem 2.3.3 Let (X, d) be a metric space, and x be a point in X. Then the following statements hold.

1. If M is a neighbourhood of x, and N is a superset of M, then N is a neighbourhood of x.
2. A finite intersection of neighbourhoods of x is a neighbourhood of x.

Proof. 1. Since M is a neighbourhood of x, there exists an open sphere $S(x, \varepsilon)$ which is contained in M. Since $N \supseteq M$,

$$S(x, \varepsilon) \subseteq M \subseteq N.$$

So N is a neighbourhood of x.

2. Let $\{N_i\}$ be a finite family of neighbourhoods of x. If this family $\{N_i\}$ is empty, then its intersection is X which is a neighbourhood of x. In the otherwise case, let the family $\{N_i\}$ contains n elements

$$N_1, N_2, \ldots, N_n.$$

Let $N = \bigcap_{i=1}^{n} N_i$. Since for each $i = 1, 2, \ldots, n$, N_i is a neighbourhood of x, there exists a real number $\varepsilon_i > 0$ such that

$$S(x, \varepsilon_i) \subseteq N_i.$$

Taking $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$, it follows that

$$S(x, \varepsilon) \subseteq S(x, \varepsilon_i) \subseteq N_i \text{ for all } i = 1, 2, \ldots, n.$$

So that

$$S(x, \varepsilon) \subseteq \bigcap_{i=1}^{n} N_i = N.$$

Thus, N is a neighbourood of x.

An arbitrary intersection of neighbourhoods of a point may not be a neighbourhood of that point. For instance, in $\mathbb{R}^2$, we consider the following sets.

$$N_n = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1/n^2\}, \text{ where } n \in Z_+.$$

For each $n \in \mathbb{Z}_+$, the set N_n is a neighbourhood of the point $(0, 0)$. The intersection $\bigcap_{n=1}^{\infty} N_n$ of the above collection is the singleton $\{(0, 0)\}$ which is not a neighbourhood of the point $(0, 0)$.

The following theorem gives an equivalent condition for a set to be neighbourhood of a point.

Theorem 2.3.4 Let (X, d) be a metric space, and $N \subseteq X$. Then N is a neighbourhood of a point $p \in X$ if and only if there exists a real number $\varepsilon > 0$ such that for some $x \in X$, $p \in S(x, \varepsilon) \subseteq N$.

Proof. If N is a neighbourhood of $p \in X$, then there exists a real number $\varepsilon > 0$ such that $p \in S(p, \varepsilon) \subseteq N$, and this is the necessity part. Conversely, since $p \in S(x, \varepsilon)$, then taking $\varepsilon_1 = \varepsilon - d(x, p)$, it follows that $S(p, \varepsilon_1) \subseteq S(x, \varepsilon) \subseteq N$. Thus, N is a neighbourhood of p.

Remark 2.3.5 In a metric space, the intersection of all neighbourhoods of a point x is the singleton $\{x\}$. To see this, consider a point x in a metric space (x, d). Let N be the

collection of all neighbourhoods of x. Let N be the intersection of all neighbourhoods of x in N. Let y be any point in X different from x so that $d(x, y) > 0$. Let $\varepsilon = d(x, y)/2$. Then the open sphere $S(x, \varepsilon)$ is a neighbourhood of x which does not contain y. So y is not in N. The point x is in N since it is in every neighbourhood in N.

Definition 2.3.6 Let (X, d) be a metric space. Then a set $N \subseteq X$ is said to be a neighbourhood of a set $A \subseteq X$ if N is a neighbourhood of every point of A. A fundamental system of neighbourhoods of A is a set $\mathcal{B}$ of neighbourhoods of A such that each neighbourhood of A contains a member of $\mathcal{B}$. If $A = \{x\}$, we speak of fundamental system of neighbourhoods of x instead of the set $\{x\}$.

Theorem 2.3.7 In any metric space, the set $\{S(x, 1/n) : n \in \mathbb{Z}_+\}$ is a fundamental system of neighbourhoods of x.

2.4 Bounded Sets

In order to represent the concept of "arbitrarily large number" and "arbitrarily small number" in a simplified manner, we need to extend the real number system by the symbols

$$-\infty(\text{minus infinity}),\ +\infty(\text{plus infinity}).$$

(The symbol $+\infty$ is usually written as ∞ without plus sign.) The resulting number system is called the extended real number system $\widetilde{\mathbb{R}}$. That is, $\widetilde{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. A number in $\widetilde{\mathbb{R}}$ is called an extended real number. Thus, an extended real number is either a real number or one of the symbols $-\infty$ and $+\infty$. By definition

$$-\infty < +\infty;$$

and if a is any real numbers, then

$$-\infty < a < +\infty.$$

The symbols $-\infty$ and $+\infty$ add essentially nothing to our understanding of real numbers. However, it may be noted that when extended real number system is assigned a metric, it may acquire characteristics which are quite distinct from that of the real line. The extended real number system when substituted for the real number system provides the notational convenience, and thus a simplification in language results. For instance, for any real number a the set

$$\{x \in \mathbb{R} : x \geq a\}$$

is denoted in the extended real number system $\mathbb{R}$ by the interval $[a, \infty[$. Similarly

$$\begin{aligned}]a, \infty[&= \{x \in \mathbb{R} : x > a\} \\ ]-\infty, a[&= \{x \in \mathbb{R} : x < a\} \\ ]-\infty, a] &= \{x \in \mathbb{R} : x \leq a\}. \end{aligned}$$

The set of real numbers $\mathbb{R}$ itself may be denoted by the open interval $]-\infty, \infty[$. The definitions of various intervals studied in $\mathbb{R}$ remain the same in $\widetilde{\mathbb{R}}$. Thus, if a, b are extended

real numbers different from both $+\infty$ and $-\infty$, then

$$\begin{aligned}
[a,b] &= \{x \in \mathbb{R} : a \le x \le b\}, \quad \text{where } a \le b,\\
[a,b[&= \{x \in \mathbb{R} : a \le x < b\}, \quad \text{where } a < b,\\
]a,b] &= \{x \in \mathbb{R} : a < x \le b\}, \quad \text{where } a < b,\\
]a,b[&= \{x \in \mathbb{R} : a < x < b\}, \quad \text{where } a < b.
\end{aligned}$$

Now let us recall the definitions of upper, lower bounds, least upper bound (l.u.b) or supremum (sup) and infimum (inf) of a subset of a partially ordered set. The real number system $\mathbb{R}$ is a partially ordered set with the usual ordering $\le$. This ordering of $\mathbb{R}$ can be extended to obtain the partial ordering of $\widetilde{\mathbb{R}}$. Specifically, we have the following . Let $a, b \in \widetilde{\mathbb{R}}$. Then the binary relation $\le_1$ on $\widetilde{\mathbb{R}}$ defined by

$$\begin{aligned}
a \le_1 b \quad &\text{if } a, b \in \mathbb{R}, \text{ and } a \le b,\\
a \le_1 b \quad &\text{if } a = -\infty, \text{ or } b = +\infty,
\end{aligned}$$

is a partial order on $\widetilde{\mathbb{R}}$.

(When there is no confusion, we denote $\le_1$ by $\le$.) Thus, $+\infty$ is the greatest element of $\widetilde{\mathbb{R}}$, and $-\infty$ is the least element of $\widetilde{\mathbb{R}}$. If A is any subset of $\mathbb{R}$ which is not bounded from above, then sup A does not exist in $\mathbb{R}$. However, when A is considered as a subset of $\widetilde{\mathbb{R}}$, sup A is $+\infty$ as any subset of $\widetilde{\mathbb{R}}$ is bounded from above. Similarly, if A is any subset of $\mathbb{R}$ which is not bounded from below, then inf A does not exist in $\mathbb{R}$ while, as a subset of $\widetilde{\mathbb{R}}$, inf $A = -\infty$. Again, in $\widetilde{\mathbb{R}}$ if A is empty, then

$$\sup A = -\infty \text{ and } \inf A = +\infty.$$

Thus, in extended real number system $\widetilde{\mathbb{R}}$, for an arbitrary subset A, it is possible to assign symbols from $\widetilde{\mathbb{R}}$ to the sup A and inf A.

As for algebraic operations in $\widetilde{\mathbb{R}}$, the corresponding algebraic operations in $\mathbb{R}$ are extended to include the following definitions. For convenience, let us call an extended real number a a finite extended real number if $a \in \mathbb{R}$. We define

$$\begin{aligned}
&a + (\pm\infty) = (\pm\infty) + a = \pm\infty \qquad \text{if } a \text{ is a finite extended real number,}\\
&(\pm\infty) + (\pm\infty) = \pm\infty,\\
&a - (\pm\infty) = (\mp\infty), \qquad \pm\infty - a = \pm\infty \qquad \text{if } a \text{ is a finite extended real number,}\\
&\pm\infty - (\mp\infty) = \pm\infty.
\end{aligned}$$

The expressions $\pm\infty + (\mp\infty)$ and $\pm\infty - \pm\infty$ remain undefined. For the operation of multiplication, we include the following definitions

$$a(+\infty) = (+\infty)a = \begin{cases} +\infty & \text{if } a > 0,\\ -\infty & \text{if } a < 0, \end{cases}$$

$$a(-\infty) = (-\infty)a = \begin{cases} +\infty & \text{if } a < 0,\\ -\infty & \text{if } a > 0, \end{cases}$$

where a is a finite extended real number. Likewise

$$0(+\infty) = (+\infty)0, \text{ and } (\pm\infty)(\pm\infty) = +\infty, \ (\pm\infty)(\mp\infty) = -\infty.$$

The extended real number system is used in extending the notion of bounded sets of real numbers to arbitrary subsets of metric spaces.

Definition 2.4.1 Let (X, d) be a metric space, and A be a subset of X. The diameter of A, written diam (A), is defined as

$$\text{diam}(A) = \sup\{d(x, y) : x, y \in A\}.$$

The diameter of A is said to be finite if diam (A) is a finite extended real number. The diameter of A is said to be infinite if it is not finite. A is said to be bounded if the diameter of A is finite.

According to the above definition, the diameter of the empty set $\emptyset$ in any metric space is infinite since diam $(\emptyset) = -\infty$. A set A in a metric space may be called unbounded if it is not bounded. If (X, d) is a metric space, and Z is any set, then a mapping $f : Z \to X$ is said to be bounded if its range set is bounded. Thus, a metric is bounded if its range set is bounded subset of the real line; when d is a bounded mapping, the metric space (X, d) is said to be bounded metric space. If (X, d) is a metric space, then the metric $\rho(x, y) = \min\{1, d(x, y)\}$ is called the standard bounded metric on X.

Sometimes we need to consider the distance of a point from a set and the distance between two sets. These are defined as follows:

Let (X, d) be a metric space, x be a point in X, and $A, B \subseteq X$. Then the distance of x from A, written $d(x, A)$, is defined by

$$d(x, A) = \inf\{d(x, y) : y \in A\}.$$

The distance between the sets A and B, denoted by $d(A, B)$, is defined by

$$d(A, B) = \inf\{d(x, y) : x \in A \text{ and } y \in B\}.$$

Examples 2.4.2

1. The intervals $[a, b]$, $]a, b]$, $[a, b[$ and $]a, b[$ are bounded sets in the real line $\mathbb{R}$. The set $A = \{1/n : n \in \mathbb{Z}_+\}$ is also a bounded set in $\mathbb{R}$. The set of rational number is an unbounded subset of $\mathbb{R}$.

2. Any non-empty subset of the discrete metric space is a bounded set. Thus, the discrete metric space itself is bounded.

3. The metric space of Example 2.1.15 is a bounded metric space, and as already remarked d is a bounded metric.

4. Any open sphere of radius $\varepsilon > 0$ in the n-dimensional Euclidean space $\mathbb{R}^n$ is bounded set with diameter 2ε.

5. Any open sphere in a metric space is a bounded set with diameter less than or equal to twice the radius.

6. Any non-empty finite subset in any metric space is a bounded set.

7. The set $\mathbb{Q} \times \mathbb{Q}$ in $\mathbb{R}^2$ is not bounded.

8. Let (X, d) be a metric space, and $A \subseteq X$. Let $D_r(A)$ be the set of all those points y of X such that $d(y, A) < r (r > 0)$, that is, $D_r(A)$ is the set of points of X which are within r distance from A. Then $D_r(A) = \bigcup\{S(x, r) : x \in A\}$. For, if $y \in D_r(A)$, then $\beta = \inf\{d(x, y) : x \in A\} < r$. Let $\varepsilon = r - \beta$. Then there exists a point $x \in A$ such that $d(x, y) < \beta + \varepsilon = \beta + r - \beta = r$ so that $y \in S(x, r)$. Conversely, if $y \in S(x, r)$ for some $x \in A$, then $d(x, y) < r$. That is, $\inf\{d(x, y) : x \in A\} < r$.

9. For any non-empty subset A of (X, d), and every point $x \in X$, $d(x, A) = d(\{x\}, A)$.

10. For any non-empty sets A and B of (X, d),

$$d(A, B) = d(B, A) = \inf\{d(x, B) : x \in A\}.$$

Theorem 2.4.3 Let (X, d) be a metric space, and A be a non-empty subset of X. Then A is bounded if and only if there exists a real number $\varepsilon > 0$ such that for some $x \in X$, $A \subseteq S(x, \varepsilon)$.

Proof. The proof is left to the reader.

Theorem 2.4.4 The union of two bounded sets is bounded.

Proof. Let (X, d) be a metric space, and A and B be two bounded subsets of X. Let $a \in A$, and $b \in B$. Let x, y be any points in $A \cup B$. Then x and y are both in A or in B, or x is in A, and y is in B. Then either $d(x, y) \leq \text{diam}(A)$, or $d(x, y) \leq \text{diam}(B)$, or in the third case $d(x, y) \leq d(x, a) + d(a, b) + d(b, y)$ which gives $\text{diam}(A \cup B) \leq d(a, b) + \text{diam}(A) + \text{diam}(B)$. Therefore,

$$\text{diam}(A \cup B) \leq d(A, B) + \text{diam}(A) + \text{diam}(B).$$

It is easy to see that if a set A in a metric space is bounded, then for any $x_0 \in X$, $d(x, x_0) \leq d(x_0, A) + \text{diam}(A)$ for all $x \in A$.

The following lemma is very useful.

Lemma 2.4.5 Let (X, d) be a metric space, and A be any non-empty subset of X. Then $|d(x, A) - d(y, A| \leq d(x, y)$ for all x, y in X.

Proof. For any point z in A, $d(x, z) \leq d(x, y) + d(y, z)$, and so

$$\begin{aligned} d(x, A) &= \inf\{d(x, z) : z \in A\} \\ &\leq \inf\{d(x, y) + d(y, z) : z \in A\} \\ &= d(x, y) + \inf\{d(y, z) : z \in A\} \\ &= d(x, y) + d(y, A). \end{aligned}$$

There are several ways in which a metric space can be associated with other metric spaces. One of them is the following.

Example 2.4.6 Let (X, d) be a metric space, and Y be a non-empty set. Let $\mathbb{B}(Y, X)$ be the set of all bounded mappings from Y into X. Let ϕ and ψ be any two element in $\mathbb{B}(Y, X)$. We define a mapping $f : Y \to \mathbb{R}$ by $f(y) = d(\phi(y), \psi(y))$. Then f is a bounded mapping. For, since ϕ and ψ are bounded, there exists real numbers k_1 and k_2 which are diameters of

$\phi(Y)$ and $\psi(Y)$, respectively. Let us fix a point $y_0 \in Y$. For any $y \in Y$,

$$\begin{aligned} f(y) &= d(\phi(y), \psi(y)) \\ &\leq d(\phi(y), \phi(y_0)) + d(\phi(y_0), \psi(y_0)) + d(\psi(y_0)), \psi(y)) \\ &\leq k_1 + d(\phi(y_0), \psi(y_0)) + k_2. \end{aligned}$$

This means that $f(Y)$ is bounded from above. In view of boundedness of f, we define a mapping $\rho : \mathbb{B}(Y, X) \times \mathbb{B}(Y, X) \to \mathbb{R}$ by setting

$$\rho(\phi, \psi) = \sup\{f(y) : y \in Y\}.$$

The mapping ρ is indeed a metric on $\mathbb{B}(Y, X)$. We prove here only the triangle inequality. For this, let ϕ, ψ and ξ be any three elements in $\mathbb{B}(Y, X)$. For $y \in Y$, we have

$$\begin{aligned} d(\phi(y), \xi(y)) &\leq d(\phi(y), \psi(y)) + d(\psi(y), \xi(y)) \\ &\leq \rho(\phi, \psi) + \rho(\psi, \xi). \end{aligned}$$

Therefore,

$$\rho(\phi, \xi) \leq \rho(\phi, \psi) + \rho(\psi, \xi).$$

Example 2.4.7 Let (X, d) be a metric space. Let $\mathbb{B}_0$ be the set of all bounded subsets of X. Let $\tau : \mathbb{B}_0 \times \mathbb{B}_0 \to \mathbb{R}$ be a mapping defined by

$$\tau(A, B) = \sup\{d(x, B) \, : \, x \in A\}.$$

Then τ is non-negative. Also $A = B$ implies that $\tau(A, B) = 0$. But $\tau(A, B) = 0$ does not imply that $A = B$. For example, in any metric space (X, d), if A is a non-empty and proper subset of a subset B of X, then $\tau(A, B) = 0$, and $A \neq B$. Further, τ is not symmetric; for instance, in the real line $\mathbb{R}$, take $A = \{1, 2\}$, $B = \{1, 2, 3\}$ in $\mathbb{B}_0$. Then $\tau(A, B) = 0$, while $\tau(B, A) = 1$. But τ satisfies the triangle inequality: Let A, B and C be any elements of $\mathbb{B}_0$. Let $r > 0$ be a real number such that $\tau(A, B) < r$. Then $A \subseteq D_r(B)$. Conversely, if $A \subseteq D_r(B)$, then $\tau(A, B) \leq r$. Thus, if $\tau(A, B) < r$, and $\tau(B, C)$ is less than a real number s, then $A \subseteq D_r(B)$, $B \subseteq D_s(C)$. Then $A \subseteq D_{r+s}(C)$, and so $\tau(A, C) \leq r + s$. Now letting r and s tend downward to $\tau(A, B)$ and $\tau(B, C)$, respectively, the desired result is obtained.

EXERCISE

1. Show that arbitrary union of bounded sets in a metric space may not be a bounded set.

2. Let A and B be two non-empty subsets of a metric space. Prove that if $A \cap B$ is non-empty, then $\text{diam}(A \cup B) \leq \text{diam}(A) + \text{diam}(B)$.

3. Let (X, d) be a metric space, x be a point in X, and A be a non-empty subset of X. Prove that $d(x, A) = 0$ if and only if every neighbourhood of x contains a point of A.

4. Prove that the function $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ defined by

$$d(x, y) = \frac{2|x - y|}{\left(\sqrt{1 + |x|^2}\sqrt{1 + |y|^2}\right)}$$

is a metric on $\mathbb{C}$. Also show that the function $\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined as $\rho(x, y) = |x/(1 + \sqrt{1 + x^2}) - y/(1 + \sqrt{1 + y^2})|$ is a metric on $\mathbb{R}$.

If $\rho(x, \infty)$, $\rho(x, -\infty)$ and $\rho(\infty, -\infty)$ are defined as limits:

$$\rho(x, \infty) = \lim_{y \to \infty} \rho(x, y),$$

$$\rho(x, -\infty) = \lim_{y \to -\infty} \rho(x, y),$$

and

$$\rho(\infty, -\infty) = \lim_{(x,y) \to (\infty, -\infty)} \rho(x, y),$$

is the function ρ a metric on the extended real number system $\widetilde{\mathbb{R}}$?

2.5 Open Sets

The notion of an open set is at the heart of modern mathematics.

Definition 2.5.1 A subset of a metric space is said to be an open set if it is a neighbourhood of each of its points.

Examples 2.5.2

1. In the real line $\mathbb{R}$, an open interval $]a, b[$, being a neighbourhood of each of its points, is an open set. The interval $[0, 1[$ is not an open set as it is not a neighbourhood of the point 0. Any closed interval $[a, b]$ is not an open set. The $\{1/n : n \in \mathbb{Z}\}$ is not an open set. The set of rational numbers is not an open set as it is not a neighbourhood of any of its points.

2. In the two dimensional Euclidean plane $\mathbb{R}^2$, the following sets are open:

   (i) $\{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 < r$, where r is a positive real number$\}$,
   (ii) $\{(x, y) \in \mathbb{R}^2 : c < y < d$, and x is arbitrary, where c and d are real numbers$\}$,
   (iii) $\{(x, y) \in \mathbb{R}^2 : c < x < d$, and y is arbitrary, where c and d are real numbers$\}$,
   (iv) $\{(x, y) \in \mathbb{R}^2 : a < x < b$, and $c < y < d$, where a, b, c and d are real numbers$\}$.

   The following sets in $\mathbb{R}^2$ are not open sets:

   (i) $\{(0, 0)\}$,
   (ii) $\{(x, y)\} \in \mathbb{R}^2 : [x^2 + y^2]^{1/2} \leq 1\}$,
   (iii) $\{(x, 2x)\} \in \mathbb{R}^2 : x$ is a real number$\}$,
   (iv) $\{(x, y)\} \in \mathbb{R}^2 : a \leq x \leq b$, y is arbitrary, where a, b are real number$\}$.

3. In $\mathbb{R}^3$, the set

$$\{(x, y, z) \in \mathbb{R}^3 : [x^2 + y^2 + z^2]^{1/2} < 1\}$$

is an open set.

4. In a discrete metric space (X, d), every subset is an open set. For, let G be any subset of X. Then for any point x in G, the open sphere $S(x, 1/2) = \{x\} \subseteq G$. Therefore, G is a neighbourhood of x. Since x is an arbitrary point in G, the assertion follows. *A* discrete metric space is richest in terms of open sets.

5. In the metric space $(\mathbb{R}^n, d_\infty)$ of Example 2.1.9, the set

$$G = \{x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0 \text{ for each } i = 1, 2, \ldots, n\}$$

is not an open set since it is not a neighbourhood of the n-tuple $(0, 0, \ldots, 0)$ in G.

6. In $\mathcal{C}[0, 1]$ (Example 2.1.10), the set G consisting of monotone functions in $\mathcal{C}[0, 1]$ is not a neighbourhood of the function $f \in \mathcal{C}[0, 1]$ given by

$$f(x) = 0 \text{ for all } x \in [0, 1].$$

7. A product of the form

$$]a_1, b_1[\times]a_2, b_2[\times \cdots \times]a_n, b_n[$$

in $\mathbb{R}^n$ is called a cell. Each cell in $\mathbb{R}^n$ (Example 2.1.8) is an open set.

Theorem 2.5.3 Every open sphere is an open set.

Proof. Let (X, d) be a metric space, and x be a point in X. Consider an open sphere $S(x, \varepsilon)$. We show that $S(x, \varepsilon)$ is an open set. Let y be any point in $S(x, \varepsilon)$. Let $d(x, y) = r$. Then $r \geq 0$. Also $r = d(x, y) < \varepsilon$. Let $\varepsilon_1 = \varepsilon - r$. Then $\varepsilon_1 > 0$. Consider the open sphere $S(y, \varepsilon_1)$. Let z be any point in $S(y, \varepsilon_1)$. By the triangle inequality, we have

$$\begin{aligned} d(x, z) &\leq d(x, y) + d(y, z) \\ &< r + \varepsilon_1. \end{aligned}$$

Therefore, $z \in S(x, \varepsilon)$. This means that $S(y, \varepsilon_1) \subseteq S(x, \varepsilon)$. It means that $S(x, \varepsilon)$ is an open set in X.

The following characterisation of open sets is frequently used.

Theorem 2.5.4 A set G in a metric space (X, d) is open if and only if for each point x in G, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G$.

Proof. Let us first assume that G is an open set in X. Let x be a point in G. Then G is a neighbourhood of x. So there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G$. Conversely, let x be a point in a set $G \subseteq X$. By our assumption, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G_i$. So G is a neighbourhood of x. Since x is an arbitrary point in G, the proof is complete.

Theorem 2.5.5 In a metric space (X, d) the following statements are true.

1. The empty set $\emptyset$ is an open set.

2. The set X is an open set.

3. A finite intersection of open sets in X is an open set in X.

4. Arbitrary union of open sets in X is an open set in X.

Proof. 1. Since there is no point in $\emptyset$ of which $\emptyset$ is not a neighbourhood, the empty set $\emptyset$ is an open set.

2. Since for any $x \in X$, any open sphere $S(x, \varepsilon) \subseteq X$, X is a neighbourhood of x. Since x is arbitrary, X is an open set.

3. Let $\{G_i\}_{i\in\Lambda}$ be a finite family of open sets in X. If the family $\{G_i\}_{i\in\Lambda}$ is empty, then its intersection is X which is an open set. So let us assume that $\{G_i\}_{i\in\Lambda}$ is non-empty finite family having n elements,

$$G_1, G_2, \ldots, G_n.$$

Let $G = \bigcap_{i=1}^{n} G_i$. If G is empty, then it is open by (1). So we assume that G is non-empty. Let x be any point in G. Then $x \in G_i$ for all $i = 1, 2, \ldots, n$. Since G_i is open, there exists a real number $\varepsilon_i > 0$ such that

$$S(x, \varepsilon_i) \subseteq G_i.$$

Assume that $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$. Then we have

$$S(x, \varepsilon) \subseteq S(x, \epsilon_i) \subseteq G_i \text{ for all } i = 1, 2, \ldots, n.$$

Therefore, $S(x, \varepsilon) \subseteq \bigcap_{i=1}^{n} G_i = G$.

4. Let $\mathcal{G} = \{G_i\}_{i\in\Lambda}$ be an arbitrary family of open sets in X. Let $G = \bigcup_{i\in\Lambda} G_i$. If $\mathcal{G}$ is empty, then G is empty, and thus is an open set. So we assume that $\mathcal{G}$ is non-empty. Again if $\mathcal{G}$ is the family containing the empty set only, then G is empty, and thus an open set. So we assume that $\mathcal{G}$ is the family containing non-empty sets. In this case G is non-empty. Let x be any point in G. Then there exists some $i \in \Lambda$ such that $x \in G_i$. Since G_i is an open set, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G_i$. Therefore,

$$S(x, \varepsilon) \subseteq G_i \subseteq G.$$

Since x is an arbitrary point, G is an open set.

It should be remarked that an arbitrary intersection of open sets may not be an open set. Consider the metric space $\mathbb{R}^2$, and the open sets in $\mathbb{R}^2$ given by

$$G_n = \{(x, y) \in \mathbb{R}^2 : -\frac{1}{n} < x < \frac{1}{n}, \text{ and } y \text{ is arbitrary}\}, \text{where } n \in \mathbb{Z}_+.$$

The intersection G of these open sets is given by

$$G = \{(x, y) \in R^2 : x = 0, \text{ and } y \text{ is arbitrary}\}$$

which is not an open set in $\mathbb{R}^2$.

Theorem 2.5.6 In a metric space (X, d), a set G is open if and only if G is a union of open spheres.

Proof. Let G be an open set in (X, d). Then for each x in G, there exists an open sphere $S(x, \epsilon) \subseteq G$, and G is the union of these open spheres. Conversely, let G be a subset of X such that it is a union of open spheres in X. If G is union of the empty family of open spheres, then G is empty, and so it is an open set. In the otherwise case G is non-empty.

Let x be any point in G. Then there exists an open sphere in the collection of which G is the union, say $S(y, \varepsilon)$ such that

$$x \in S(y, \varepsilon) \subseteq G.$$

Since $S(y, \varepsilon)$ is an open set, there exists a real number $\epsilon' > 0$ such that

$$x \in S(x, \varepsilon') \subseteq S(y, \varepsilon).$$

This means that $S(x, \varepsilon')$ is contained in G, which proves that G is an open set.

There are certain metric spaces which are at the foundation of advanced analysis. The properties of these individual spaces have been explored. The following theorems are attempts in that direction.

Theorem (Lindelöf) 2.5.7 If $\mathcal{G} = \{I_\alpha\}_{\alpha \in \Lambda}$ is a collection of open intervals in the real line $\mathbb{R}$, then there exists an atmost countable subset Λ' of Λ such that

$$\bigcup_{\alpha \in \Lambda'} I_\alpha = \bigcup_{\alpha \in \Lambda} I_\alpha.$$

Proof. Each $x \in I_\alpha$, $\alpha \in \Lambda$, is contained in an open interval J_i with rational end points such that $J_i \subseteq I_\alpha$. Let us denote by Λ' the set of suffixes of J_i obtained in this way. Since rationals are countable, Λ' is atmost countable. Also $\bigcup_{i \in \Lambda'} J_i = \bigcup_{\alpha \in \Lambda} I_\alpha$. For, let $x \in \bigcup_{\alpha \in \Lambda} I_\alpha$. Then $x \in I_\alpha$ for some $\alpha \in \Lambda$. So there exists an $i \in \Lambda'$ such that $x \in J_i$. Then $x \in \bigcup_{\alpha \in \Lambda} J_i$. Now for each open interval J_i in the collection $\{J_i\}_{i \in \Lambda'}$, there exists an open interval I_α in $\mathbb{G}$ such that $J_i \subseteq I_\alpha$. Let us rename the suffix α as α_i. Then the collection $\{I_{\alpha_i}\}_{i \in \Lambda'}$ is at most countable, and we have

$$\bigcup_{\alpha \in \Lambda} I_\alpha = \bigcup_{i \in \Lambda'} J_i = \bigcup_{i \in \Lambda'} I_{\alpha_i}.$$

The proof is now complete.

Theorem (Lindelöf) 2.5.8 If $\mathcal{G} = \{G_\alpha\}_{\alpha \in \Lambda}$ is a collection of open sets in $\mathbb{R}^n$, then there exists an at most countable subset Λ' of Λ such that

$$\bigcup_{i \in \Lambda'} G_i = \bigcup_{i \in \Lambda} G_i.$$

Proof. Let $x = (x_1, x_2, \ldots, x_n)$ be a point in G_α for some $\alpha \in \Lambda$. Then there exists an open sphere $S(x, \varepsilon)$ contained in G_α. Consider a "cube" T_α with sides of length $\varepsilon/\sqrt{n}$ such that $x \in T_\alpha$, and a "rectangle" J_i with rational co-ordinates for its vertices such that $x \in J_i$, and $J_i \subseteq T_\alpha$. Note that $T_\alpha \subseteq S(x, \varepsilon)$. Now the proof proceeds as in the case of $\mathbb{R}$, Theorem 2.5.7 above.

Theorem 2.5.9 Each non-empty open set G in $\mathbb{R}$ is the union of disjoint open intervals, at most countable in number.

Proof. We define an equivalence relation on G: Let $a \sim b$ if the closed interval $[a, b]$, or $[b, a]$ if $b \leq a$, is contained in G. The relation $\sim$ is reflexive since $[a, a] = \{a\} \subseteq G$ for each $a \in G$. If for $a, b \in G$, $a \sim b$, then $[a, b]$ or $[b, a] \subseteq G$. So $b \sim a$. The relation $\sim$ is obviously transitive. Thus, the relation $\sim$ partitions G into equivalence classes. Let us denote the equivalence class of an element a in G by $[a]$. We show that $[a]$ is an interval. Let b and c

$(b < c)$ be any two elements in $[a]$, and d be such that $b < d < c$. Since $b \in [a]$, $[a, b] \subseteq G$, or $[b, a] \subseteq G$. Similarly, $[a, c] \subseteq G$, or $[c, a] \subseteq G$. Assume that $a \leq b$. Then $[a, b] \subseteq G$, and $[a, c] \subseteq G$ which means that $[a, d]$ is contained in G. So $a \sim d$ in this case. Now assume that $b < a$. Then $[b, a] \subseteq G$. If $a \leq c$, then $[a, c] \subseteq G$. It follows in this case that either $[a, d]$ or $[d, a]$ is contained in G. If $c < a$, then $[c, a] \subseteq G$. The fact that $[b, a] \subseteq G$ together with $[c, a] \subseteq G$ will imply that $[d, a] \subseteq G$. So $a \sim d$, and $d \in [a]$. This proves that $[a]$ is an interval. Now we show that $[a]$ is open. Let b be any element in $[a]$. Since G is open, there exists a real number $\varepsilon > 0$ such that the open interval $]b - \varepsilon, b + \varepsilon[$ is contained in G. But this means that the interval $]b - \varepsilon, b + \varepsilon[$ is contained in $[a]$. Therefore, G is the union of disjoint open intervals. Then by Theorem 2.5.7, G is the union of at most countable disjoint open intervals.

The following theorem is known as Hausdorff property in metric spaces.

Theorem 2.5.10 In a metric space (X, d), for any distinct points x and y, there exist disjoint open sets U and V such that $x \in U$, and $y \in V$.

Proof. Since x and y are distinct, $d(x, y) > 0$. Let $\varepsilon = d(x, y)/2$. Then it is easily seen that $S(x, \varepsilon)$ and $S(y, \varepsilon)$ are disjoint open spheres, and hence open sets such that $x \in S(x, \varepsilon)$, and $y \in S(y, \varepsilon)$.

Theorem 2.5.11 Let (X, d) be a metric space, and A be a subset of X. For any real number $\varepsilon > 0$, the set $D_\varepsilon(A) = \{x \in X : d(x, A) < \varepsilon\}$ is an open set in X.

Proof. Let $x \in D_\varepsilon(A)$. Then $d(x, A) < \varepsilon$. Let $\varepsilon_1 = \varepsilon - d(x, A)$. Let y be any point in $S(x, \varepsilon_1)$. Then $d(x, y) < \varepsilon_1$, that is, $d(x, y) < \varepsilon - d(x, A)$. By Lemma 2.4.5,

$$d(y, A) \leq d(x, y) + d(x, A) < \varepsilon.$$

Therefore, $S(x, \varepsilon_1) \subseteq D_\varepsilon(A)$, and $D_\varepsilon(A)$ is an open set.

EXERCISE

1. In the metric space $\mathcal{C}[0, 1]$ of Example 2.1.10, the set $G = \{f \in \mathcal{C}[0, 1] : f(1/2) > 5\}$ is an open set.

2. Let A be the unit circle $\{(\cos\theta, \sin\theta) : 0 \leq \theta < 2\pi\}$ in $\mathbb{R}^2$. Let $x_1 = (\cos\theta_1, \sin\theta_1)$, and $x_2 = (\cos\theta_2, \sin\theta_2)$ be in A. Define $d(x_1, x_2) = |\theta_1 - \theta_2|$. Show that d is metric on A. What are open sets in A?

3. Show that the complement of a finite set in any metric space is an open set.

2.6 Equivalent Metrics

It is possible that different metrics on a set give rise to same open sets. There are certain properties which depend only on the open sets in a metric space. If such a property is known to be true in a metric space (X, d), then it turns out that this property will be true in all metric spaces which have the same open sets as are in (X, d). Such properties are known as topological properties. We shall be identifying such properties as we proceed along in the text.

Definition 2.6.1 Let d_1 and d_2 be two metrics on a non-empty set X. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be the sets of open sets in (X, d_1) and (X, d_2), respectively. Then d_1 is said to be finer than d_2 if $\mathcal{G}_2 \subseteq \mathcal{G}_1$. Two metrics d_1 and d_2 are said to be equivalent if $\mathcal{G}_1 = \mathcal{G}_2$.

We illustrate with the following example.

Example 2.6.2 Let (X, d) be a metric space. Then the metric ρ on X as defined in Example 2.1.16 is equivalent to d. To show this, let G be an open set in (X, d), and x be a point in G. Then there exists a real number ε such that $0 < \epsilon < 1$, and $S(x, \varepsilon, d) \subseteq G$. Then $\rho(x, y) < \varepsilon$ if and only if $\min\{1, d(x, y)\} < \varepsilon$ if and only if $d(x, y) < \varepsilon$. Therefore, $S(x, \varepsilon, \rho) = S(x, \varepsilon, d) \subseteq G$. Thus, G is an open set in (X, ρ). To show that every open set in (X, ρ) is an open set in (X, d), let G^* be an open set in (X, ρ), and x be a point in G^*. Then there exists a real number ε such that $0 < \varepsilon < 1$, and $S(x, \varepsilon, \rho) \subseteq G^*$. It is easily seen that $S(x, \varepsilon, d) = S(x, \varepsilon, \rho) \subseteq G^*$. Thus, G^* is an open set in (X, d). This proves that d and ρ are equivalent metrics.

The above example is in fact a proof of the following theorem.

Theorem 2.6.3 Any metric d on a set X is equivalent to a bounded metric ρ on X in which the diameter of X does not exceed 1.

Theorem 2.6.4 Let (X, d) and $(X, d'$ be two metric spaces. Then d and d' are equivalent if and only if

(i) for each $x \in X$ and for each open sphere $S(x, \varepsilon, d)$, there exists an open sphere $S(x, \varepsilon', d')$ contained in $S(x, \varepsilon, d)$, and

(ii) for each $x \in X$ and for each open sphere $S(x, \varepsilon, d')$, there exists an open sphere $S(x, \varepsilon', d)$ contained in $S(x, \varepsilon, d')$.

Proof. First assume that d and d' are equivalent. Since an open sphere is an open set, for each x and open sphere $S(x, \varepsilon, d)$, there exists a real number $\varepsilon' > 0$ such that $S(x, \varepsilon', d') \subseteq S(x, \varepsilon, d)$. Similarly, (b) can be proved. Conversely, we assume (a) and (b), and show that d and d' are equivalent. Let G be an open set in (X, d), and x be a point in G. So there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon, d) \subseteq G$. By the assumption (a), there is a real number $\varepsilon' > 0$ such that $S(x, \varepsilon', d') \subseteq S(x, \varepsilon, d)$. Therefore, $S(x, \varepsilon', d') \subseteq G$. Thus, G is an open set in (X, d'). Similarly, using the assumption (b), it follows that any open set in (X, d') is an open set in (X, d). Thus, d and d' are equivalent.

Here is another example on similar lines.

Example 2.6.5 Let (X, d) be a metric space, and d_1 be as defined in Example 2.1.15. Then d and d_1 are equivalent. Let x be any point in X, and y be any point in $S(x, \varepsilon, d)$. Then $y \in S(x, \varepsilon, d)$ if and only if $d(x, y) < \varepsilon$, if and only if $d(x, y)/(1 + d(x, y)) < \varepsilon/(1 + \varepsilon)$ if and only if $d_1(x, y) < \varepsilon/(1+\varepsilon)$. Consequently, $S(x, \varepsilon/(1+\varepsilon), d_1) = S(x, \varepsilon, d)$. For a real number ε such that $0 < \varepsilon < 1$, a similar calculation will show that $S(x, \varepsilon, d_1) = S(x, \varepsilon/(1 - \varepsilon), d)$. Hence, by Theorem 2.6.4, d and d_1 are equivalent.

From Examples 2.6.2 and 2.6.5, it follows that d, ρ and d_1 are equivalent metrics. The following theorem gives a sufficient condition for two metrics to be equivalent.

Theorem 2.6.6 Two metric d and d^* on a non-empty set X are equivalent if

$$sd(x, y) \leq d^*(x, y) \leq td(x, y)$$

for all x, y in X and for some positive real numbers s and t.

Proof. Suppose that the real numbers s and t exist as required. For any $x \in X$, and a real number $\varepsilon > 0$

$$S(x, \varepsilon, d) \subseteq S(x, t\varepsilon, d^*).$$

Again for any $y \in X$, and a real number $\epsilon' > 0$

$$S(y, \varepsilon', d^*) \subseteq S(y, \varepsilon'/s, d).$$

Therefore, d and d^* are equivalent by Theorem 2.6.4.

The sufficient condition in the above theorem is clearly not a necessary condition.

Let us consider the real line $\mathbb{R}$, where the metric is $d(x, y) = |x - y|$. This metric d is equivalent to the metric d_1 on $\mathbb{R}$ defined by

$$d_1(x, y) = d(x, y)/(1 + d(x, y)).$$

Suppose, if possible, there exists a real number $t > 0$ such that $d(x, y) \leq td_1(x, y)$. Then $d(x, y) \leq t$, that is, $|x - y| \leq t$ for all x and y in $\mathbb{R}$ which is clearly not possible.

Recall that a sequence in a set X is a mapping $s : \mathbb{N}$ (or $\mathbb{N}_0$) $\to X$, and if $s(n) = x_n \in X$, then the sequence s is denoted in terms of its values $< x_n >$. Whenever there is a metric d defined on X, then $< x_n >$ is said to be in the metric space X. Such a sequence $< x_n >$ is said to be convergent to a point x in X if for each real number $\varepsilon > 0$, there exists a positive integer m such that $d(x_n, x) < \varepsilon$ for all $n \geq m$. (See, detailed discussion in Chapter 3).

Two metrics preserve the convergent sequences if and only the metrics are equivalent. This is our next characterisation of equivalent metrics.

Theorem 2.6.7 Two metrics d and d' on a non-empty set X are equivalent if and only if (i) every sequence $< x_n >$ in X converging to x in (x, d) converges to x in (X, d'), and (ii) every sequence $< x_n >$ in X converging to x in (X, d') converges to x in (X, d).

Proof. Assume that d and d' are equivalent. Let x be a point in X, and $\varepsilon > 0$ be a real number. By Theorem 2.6.4, there exists a real number $\delta > 0$ such that $S(x, \delta, d) \subseteq S(x, \varepsilon, d')$. If $< x_n >$ converges to $x \in X$ in (X, d), then there exists a positive integer p such that $d(x_n, x) < \delta$ for all $n \geq p$. Thus, we have

$$x_n \in S(x, \delta, d) \subseteq S(x, \varepsilon, d') \text{ for all } n \geq p.$$

That is, $< x_n >$ converges to x in (X, d'). The proof of (ii) is similar. Conversely, we assume the conditions (i) and (ii) in the theorem, and prove that d and d' are equivalent. Suppose the contrary. By Theorem 2.6.4, there is a point $x_0 \in X$, and a real number $\varepsilon > 0$ such that for all real numbers $\delta > 0$, $S(x_0, \delta, d')$ is not contained in $S(x_0, \varepsilon, d)$. In particular, for each $n \in \mathbb{Z}_+$ since $S(x_0, 1/n, d')$ is not contained in $S(x_0, \varepsilon, d)$, we can select a point $x_n \in S(x_0, 1/n, d')$ such that x_n is not in $S(x_0, \varepsilon, d)$. Now it is easily seen that the sequence $< x_n >$ converges to x_0 in (X, d'), but does not converge to x_0 in (X, d), which contradicts the assumption (ii). The proof is now complete.

The inequality in the following lemma is used to establish the equivalence of metric d_p in Example 2.1.8 and d_∞ in Example 2.1.9.

Lemma 2.6.8 For all $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{F}^n$

$$d_\infty(x, y) \leq d_p(x, y) \leq n^{1/p} d_\infty(x, y), \text{ where } 1 \leq p < \infty.$$

Proof. We have

$$\begin{aligned} d_\infty(x,y) &= \max\{|x_i - y_i| : i = 1,2,\ldots,n\} \\ &\leq \sum_{i=1}^{n} |x_i - y_i| = d_1(x,y) \leq n \max\{|x_i - y_i| : i = 1,2,\ldots,n\}. \end{aligned}$$

This means that $d_\infty(x,y) \leq d_1(x,y) \leq n d_\infty(x,y)$. Also for $1 \leq p < \infty$,

$$d_p(x,y) \geq |x_i - y_i| \text{ for all } i = 1,2,\ldots,n.$$

Therefore,

$$d_p(x,y) \geq \max\{|x_i - y_i| : i = 1,2,\ldots,n\} = d_\infty(x,y).$$

On the other hand,

$$\begin{aligned} (d_p(x,y))^p &= \sum_{i=1}^{n} |x_i - y_i|^p \leq n \max\{|x_i - y_i|^p : i = 1,2,\ldots,n\} \\ &= n[\max\{|x_i - y_i| : i = 1,2,\ldots,n\}]^p. \end{aligned}$$

The required inequality now follows.

Example 2.6.9 The metrics d_p in Example 2.1.8 and d_∞ in Example 2.1.9 are equivalent. For any point $a \in \mathbb{F}^n$ and a real number $\varepsilon > 0$, the containments

$$S(a, n^{-1/p}\varepsilon, d_\infty) \subseteq S(a, \varepsilon, d_p) \subseteq S(a, \varepsilon, d_\infty)$$

follow from Lemma 2.6.8. The equivalence of d_p and d_∞ is obvious by Theorem 2.6.4.

Example 2.6.10 Let (X,d) be a metric space. Let $h : [0,\infty[\to [0,\infty[$ be a strictly increasing function such that $h(0) = 0$, and $h(t)/t$ is decreasing. Then $h(x+y) \leq h(x) + h(y)$. The mapping $h \circ d$ is a metric on X. In particular, if $\alpha \in]0,1]$, d^α is a metric on X. Also $h \circ d$ and d are equivalent. To see this, let x be a point in X, and $\varepsilon > 0$ be a real number. Then $h(\varepsilon) > 0$, and

$$S(x, h(\varepsilon), h \circ d) \subseteq S(x, \varepsilon, d).$$

Also let δ be a real number such that $0 < \delta < \varepsilon$. Then $y \in S(x, h^{-1}(\delta), d)$ implies that $h(d(x,y)) < \delta < \varepsilon$, that is, $y \in S(x, \varepsilon, h \circ d)$. Therefore, d and $h \circ d$ are equivalent.

Let $M(X)$ be the set of all metrics on a non-empty set X. For d and d' in $M(X)$, we write $d \sim d'$ if d and d' are equivalent metrics. The relation $\sim$ is obviously an equivalence relation on $M(X)$.

EXERCISE

1. Let $\mathcal{C}[a,b]$ be the set of all real valued continuous mappings. Show that the following metrics on $\mathcal{C}[a,b]$ (See Example 2.1.10 and 2.1.11)

$$d_1(f,g) = \sup\{|f(t) - g(t)| : t \in [a,b]\}, \text{ and } d_2(f,g) = \int_a^b |f(t) - g(t)|dt$$

are not equivalent.

2. Let $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a mapping defined by

$$d(x,y) = \begin{cases} 1 + |y - x| & \text{if one and only one of the real numbers } x \text{ and } y \\ & \text{is positive,} \\ |y - x|. & \text{otherwise.} \end{cases}$$

Prove that d is a metric on $\mathbb{R}$ which is not equivalent to the usual metric on $\mathbb{R}$.

3. Let $d :]0,1[\times]0,1[\to \mathbb{R}$ be a mapping defined by

$$d(x,y) = |1/y - 1/x| \text{ for all } x, y \text{ in }]0,1[.$$

Prove that d is a metric on $]0,1[$ which is equivalent to the usual metric on $]0,1[$.

4. Let X be any finite non-empty set X. Show that any metric on X is equivalent to the discrete metric on X.

5. Let $\{(X_i, d_i)\}_{i \in \mathbb{Z}_+}$ be a family of metric spaces. Show that the mappings

$$d(x,y) = \sum_{i=1}^{\infty} d_i(x_i, y_i)/2^i(1 + d_i(x_i, y_i)),$$

and $$d^*(x,y) = \sum_{i=1}^{\infty} \min\{1, d_i(x_i, y_i)\}/2^i$$

for all $x =< x_i >$, and $y =< y_i >$ in $\prod_{i=1}^{\infty} X_i$ are metrics on $\prod_{i=1}^{\infty} X_i$, and these metrics are equivalent.

2.7 Closed Sets

The dual of the concept of an open set is the concept of a closed set.

Definition 2.7.1 A set in a metric space X is said to be a closed set if its complement in X is an open set in X.

Examples 2.7.2

1. Any closed interval $[a,b]$ is a closed set in the real line $\mathbb{R}$. Also the sets $\{x \in \mathbb{R} : x \geq a\}$ and $\{x \in \mathbb{R} : x \leq a\}$, where a is a real number, are closed sets in $\mathbb{R}$. Any finite set is a closed set in $\mathbb{R}$. The set of integers is a closed set in $\mathbb{R}$. The set $\mathbb{Q}$ of rational numbers is neither open nor a closed set in $\mathbb{R}$. The set $\{1/n : n \in \mathbb{Z}_+\}$ is not a closed set in $\mathbb{R}$.

2. Let us consider the two dimensional Euclidean plan $\mathbb{R}^2$. The sets $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ and $\{(x, mx) : m \text{ is a real number}\}$ are closed sets in $\mathbb{R}^2$. The set $\mathbb{Q} \times \mathbb{Q}$ in $\mathbb{R}^2$ is neither open nor a closed set in $\mathbb{R}^2$. The sets $\{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$, $\{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1, \text{ and } y \text{ is arbitrary}\}$, and $\{(x,y) \in \mathbb{R}^2 : 1 \leq y \leq 2, \text{ and } x \text{ is arbitrary}\}$ are closed sets in $\mathbb{R}^2$.

3. The set $\{f \in \mathcal{C}[a,b] : f(b) = 1$ is a closed set in $\mathcal{C}[a,b]\}$.

4. The set $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2z^2 \leq 1\}$ is a closed set in three dimensional Euclidean space $\mathbb{R}^3$.

5. In a metric space, every finite set is a closed set.

Theorem 2.7.3 Let (X, d) be a metric space. Then the following statements hold.

1. X and $\emptyset$ are closed sets.
2. union of a finite number of closed sets in X is a closed set in X.
3. Arbitrary intersection of closed sets in X is a closed set in X.

Proof. 1. The complement of X in X is $\emptyset$ which is an open set; so X is a closed set in X. The complement of $\emptyset$ in X is X which is an open set; so $\emptyset$ is a closed set.

2. Let $\{F_i\}_{i\in\Lambda}$ be a finite family of closed sets in X. Let $F = \bigcup_{i\in\Lambda} F_i$. If $\{F_i\}_{i\in\Lambda}$ is empty, then F is empty, and thus a closed set. So we assume that the family $\{F_i\}_{i\in\Lambda}$ is non-empty, and contains n elements $F_1, F_2, \ldots, F_n$. Then, by De Morgan's law

$$F' = \left(\bigcup_{i=1}^{n} F_i\right)' = \bigcap_{i=1}^{n} F_i',$$

where F' is the complement of F in X etc. The intersection on the right of the above equality is an open set by Theorem 2.5.5. Therefore, F is a closed set.

3. Let $\mathbb{F} = \{F_i\}_{i\in\Lambda}$ be an arbitrary family of closed sets in X and $F = \bigcap_{i\in\Lambda} F_i$. If $\mathbb{F}$ is empty, then F is equal to X. Since X is a closed set, F is closed. If $\mathbb{F}$ is non-empty, then by De Morgan's law

$$F' = \left(\bigcap_{i\in\Lambda} F_i\right)' = \bigcup_{i\in\Lambda} F_i'.$$

The set on the right in the above equality is open by Theorem 2.5.5, and so F is a closed set.

Arbitrary union of closed sets may not be a closed set. We consider in $\mathbb{R}^2$ the following sets.

$$F_n = \{(x, y) \in \mathbb{R}^2 : (x^2 + y^2)^{1/2} \leq n/(n + 1)\} \text{ for all } n \in \mathbb{Z}_+.$$

The union of these sets is

$$\bigcup_{n=1}^{\infty} F_n = \{(x, y) \in \mathbb{R}^2 : (x^2 + y^2)^{1/2} < 1\}$$

which is not a closed set in $\mathbb{R}^2$.

The following theorem is easy to prove.

Theorem 2.7.4 In a metric space (X, d), every closed sphere is a closed set.

Example 2.7.5 Let (X, d) be a metric space, and $A, B \subseteq X$. If A is closed, and B is open in X, then $A - B$ is a closed set in X. To see this, let x be any point in $(A - B)'$, the complement of $A - B$ in X. If x is in A', then there is a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq A'$ so that $S(x, \varepsilon) \subseteq (A - B)'$. If x is in $A \cap B$, then there exists a real number

$\varepsilon' > 0$ such that $S(x, \varepsilon') \subseteq B \subseteq (A - B)'$. Thus, $(A - B)'$ is an open set in X, and $A - B$ is a closed set in X. Also it can be easily seen that $B - A$ is an open set in X.

For the illustration of the following example, the reader is referred to Example 2.1.19.

Examples 2.7.6 Let (X, d) be a metric space satisfying the ultrametric inequality

$$d(x, z) \leq \max\{d(x, y), d(y, z)\} \text{ for all } x, y \text{ and } z \text{ in } X.$$

This space has several interesting properties. We observe the following.

1. Let us assume that $d(x, y) \neq d(y, z)$. If $d(x, y) < d(y, z)$, then it follows that $d(x, z) \leq d(y, z)$. The inequality $d(y, z) \leq \max\{d(y, x), d(x, z)\}$ gives $d(y, z) \leq d(x, z)$ so that $d(x, z) = d(y, z)$. Thus, if $d(x, y) \neq d(y, z)$, then $d(x, z) = \max\{d(x, y), d(y, z)\}$.

2. Let y be any point in X, and $\varepsilon > 0$ be a real number. Then $S(y, \varepsilon)$ is a closed set in X. To see this, let x be any point in $X - S(y, \varepsilon)$. Then $d(x, y) \geq \varepsilon$. Let z be any point in $S(y, \varepsilon)$. Then $d(y, z) < \varepsilon$. Therefore, $d(x, z) = \max\{d(x, y), d(y, z)\} = d(x, y)$. Hence, any open sphere $S(x, \varepsilon)$ is disjoint from $S(y, \varepsilon)$, and $S(y, \varepsilon)$ is a closed set.

3. Any closed sphere in (X, d) is an open set. Assume that $S[y, \varepsilon]$ is a closed sphere in X, and z is a point in $S[y, \varepsilon]$. It can be easily seen that for any real number ε', $0 < \varepsilon' < \varepsilon$, the open sphere $S(z, \varepsilon')$ is contained in $S[y, \varepsilon]$. Also if $x \in S[y, \varepsilon]$, then $S[x, \varepsilon] = S[y, \varepsilon]$.

4. If two (closed) open spheres have a common point, then one of them is contained in the other. Let z be a point in $S(x, \varepsilon) \cap S(y, \varepsilon')$, where $\varepsilon \leq \varepsilon'$. Let u be a point in $S(x, \varepsilon)$. Then, by the inequalities $d(u, z) \leq \max\{d(u, x), d(x, z)\}$, and $d(u, y) \leq \max\{d(u, z), d(z, y)\}$, it follows that $u \in S(y, \varepsilon')$. Again let $z \in S[x, \varepsilon] \cap S[y, \varepsilon']$, and $\varepsilon \leq \varepsilon'$. Then similarly as above it can be shown that $S[x, \varepsilon] \subseteq S[y, \varepsilon']$.

5. The distance between two disjoint open spheres of radius $\varepsilon > 0$ contained in a closed sphere of radius ε, is equal to ε. Assume that $S(x, \varepsilon) \subseteq S[y, \varepsilon]$ and $S(x', \varepsilon) \subseteq S[y, \varepsilon]$, and $S(x, \varepsilon) \cap S(x', \varepsilon) = \emptyset$. Let $u \in S(x, \varepsilon)$, and $v \in S(x', \varepsilon)$. Then $d(u, v) \geq \varepsilon$. Now, the three inequalities

$$d(u, v) \leq \max\{d(u, x), d(x, v)\},$$
$$d(u, x) \leq \max\{d(u, y), d(y, x)\},$$
$$d(x, v) \leq \max\{d(x, y), d(y, v)\},$$

show that $d(u, v) \leq \varepsilon$, and the result follows.

EXERCISE

1. Prove that the set of irrational numbers is not the union of a countable number of closed sets.

2. For any metric space (X, d), any real number $r > 0$ and any non-empty subset A of X, the set $D'_r(A) = \{x \in X : d(x, A) \leq r\}$ is a closed set in X.

3. Let (X, d) be a metric space, and $\{A_\lambda\}_{\lambda\in\Lambda}$ be a family of closed sets in X with the property that there exists a real number $\delta > 0$ such that $d(A_\lambda, A_{\lambda'}) \geq \delta$ for all $\lambda \neq \lambda'$. Prove that the union $\bigcup\{A_\lambda : \lambda \in \Lambda\}$ is a closed set in X.

4. If $A = \{2, 3, 4, \ldots\}$ and $B = \{2 + \frac{1}{2}, 3 + \frac{1}{3}, 4 + \frac{1}{4}, \ldots\}$ are subsets of the real line $\mathbb{R}$. Show that $d(A, B) = 0$. (The distance between two disjoint closed sets may be zero.)

5. Let (X, d) be a metric space. Prove that

   (a) $\bigcap\{F : p \in F, \text{ and } F \text{ is closed}\} = \{p\}$.
   (b) $\bigcap\{G : p \in G, \text{ and } G \text{ is open}\} = \{p\}$.

6. Show that every closed subset of a metric space is the intersection of a countable number of open sets.

2.8 Limit Point and Closure of a Set

Definition 2.8.1 Let (X, d) be a metric space, and A be a subset of X. A point $p \in X$ is said to be a limit point of A if every neighbourhood of p contains a point of A other than p.

The set of all limit points of A is called the derived set of A, denoted by $D(A)$.

The following theorem gives equivalent condition for a point to be a limit point of a set.

Theorem 2.8.2 In a metric space (X, d), a point $p \in X$ is a limit point of a set $A \subseteq X$ if and only if for each real number $\varepsilon > 0$, the open sphere $S(p, \varepsilon)$ contains a point of A other than p.

Proof. The necessity part is obvious since every open sphere centered at p is a neighbourhood of p, and so by our assumption contains a point of A other than p.

Conversely, if N is a neighbourhood of p, then there exists a real number $\varepsilon > 0$ such that $S(p, \varepsilon) \subseteq N$. By our assumption, $S(p, \varepsilon)$ and hence N contains a point of A other than p.

Examples 2.8.3

1. Let us consider the real line $\mathbb{R}$. Let $A =]0, 1[$. Any point $p \in \mathbb{R}$ such that $0 \leq p \leq 1$ is a limit point of A since for any positive real number ε, the open sphere $S(p, \varepsilon)$ which is equal to the open interval $]p - \varepsilon, p + \varepsilon[$ contains a point of A other than p. If $p < 0$, and $\varepsilon = |p|$, then the open sphere $]p - \varepsilon, p + \varepsilon[$ contains no point of A. Similarly, any point $p > 1$ is not a limit point of A. Thus, $D(A) = [0, 1]$.

2. Every real number is a limit point of the set of rational numbers $\mathbb{Q}$ in the real line $\mathbb{R}$. Thus, $D(\mathbb{Q}) = \mathbb{R}$.

3. In the real line $\mathbb{R}$, the set $A = \{1/n : n \in \mathbb{Z}_+\}$ has zero as its only limit point. Thus $D(A) = \{0\}$.

4. In $\mathbb{R}^2$, each point of the circle $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is a limit point of A, and no point of $\mathbb{R}^2 - A$ is a limit point of A. Similarly, for any straight line in $\mathbb{R}^2$ the set of limit points is the line itself. The set of limit points of $\mathbb{Q} \times \mathbb{Q}$ in $\mathbb{R}^2$ coincides with $\mathbb{R}^2$. Every point p in $\mathbb{R}^2$ is a limit point of $\mathbb{Q} \times \mathbb{Q}$ since every open sphere $S(p, \varepsilon)$ contains a point of $\mathbb{Q} \times \mathbb{Q}$ other than p. The set $A = \{(0, 1/n) \in \mathbb{R}^2 : n \in \mathbb{Z}_+\}$ has $(0, 0)$ as its only limit point.

5. In a discrete metric space (X, d), no point of X is a limit point of any subset of X.

6. In a metric space (X, d), no point of X can be a limit point of any finite subset of X. For, if the finite set is empty, then no point of X can be a limit point of the empty set. If the finite set is $A = \{x_1, x_2, \ldots, x_n\}$ having n elements of X, and p be any point in X, then by taking $\varepsilon = \min\{d(p, x_i) : x_i \in A, \text{ and } p \neq x_i\}$, it is seen that the open sphere $S(p, \varepsilon)$ contains no point of A other than p.

Here is another equivalent condition for a point to be a limit point.

Theorem 2.8.4 Let (X, d) be a metric space, and $A \subseteq X$. A point $p \in X$ is a limit point of A if and only if every neighbourhood of p contains infinitely many points of A.

Proof. Suppose that a point p in X is a limit point of A. Let N be a neighbourhood of p which contains only a finite number of points of A. Let the points in $N \cap A - \{p\}$ be $x_1, x_2, \ldots, x_n$. We define $\varepsilon = \min\{d(p, x_i) : i = 1, 2, \ldots, n\}$. Then $S(p, \varepsilon) \cap N$ is a neighbourhood of p which does not contain any point of A other than p. This means that p is not a limit point of A, and this contradicts our hypothesis. Thus, every neighbourhood of p contains infinitely many points of A.

Conversely, if every neighbourhood of p contains infinitely many points of A, then every neighbourhood of p contains a point of A other than p. Thus, p is a limit point of A.

Theorem 2.8.5 In a metric space (X, d), a subset A of X is closed if and only if A contains all its limit points.

Proof. If A is closed, then its complement $X - A$ is an open set in X. Thus, for any point $p \in X - A$, $X - A$ is a neighbourhood of p which does not contain any point of A. Therefore, any point in $X - A$ is not a limit point of A. Thus, A contains all its limit points. Conversely, assume that A contains all its limit points. Let p be a point in $X - A$. Then p is not a limit point of A, and so there exists an open sphere $S(p, \varepsilon)$ which contains no point of A. That is, $S(p, \varepsilon) \subseteq X - A$. Since p is an arbitrary point, $X - A$ is an open set, and the proof is complete.

Corollary 2.8.6 In a metric space every finite set is a closed set.

Proof. Let A be a finite subset of a metric space (X, d). By Theorem 2.8.3, no point of X can be a limit point of A. So there is no point in $X - A$ which is a limit point of A. By Theorem 2.8.4, A is a closed set.

The following theorem gives some useful results about limit points.

Theorem 2.8.7 Let (X, d) be a metric space, and A and B be two subsets of X. Then

(i) $D(\emptyset) = \emptyset$,

(ii) $A \subseteq B$ implies that $D(A) \subseteq D(B)$,

(iii) $D(A \cup B) = D(A) \cup D(B)$,

(iv) $x \in D(A)$ implies that $x \in D(A - \{x\})$.

Proof. (i) Let p be any point in X, and N be a neighbourhood of p. Then $N \cap \emptyset = \emptyset$. So p can not be a limit point of $\emptyset$. Therefore, $D(\emptyset) = \emptyset$.

(ii) Let $p \in X$ be a limit point of A. Then every neighbourhood N of p contains a point of A and hence of B other than p. So p is a limit point of B. Thus, $D(A) \subseteq D(B)$.

(iii) Since $A \subseteq A \cup B$, and $B \subseteq A \cup B$, by (2), it follows that $D(A) \cup D(B) \subseteq D(A \cup B)$. To show the reverse containment, suppose that $x \notin D(A) \cup D(B)$. This means that x is neither

a limit point of A nor a limit point of B. Then there exists a neighbourhood M of x which does not contain any point of A other than x. Similarly, there is a neighbourhood N of x which does not contain any point of B other than x. Therefore, $M \cap N$ is a neighbourhood of x which does not contain any point of $A \cup B$ other than x. Thus, x is not a limit point of $A \cup B$. That is $x \notin D(A \cup B)$.

(iv) Let $x \in D(A)$, and N be a neighbourhood of x. Therefore, N contains a point of $A - \{x\}$ other than x. Thus, x is a limit point of $A - \{x\}$, that is, $x \in D(A - \{x\})$.

Theorem 2.8.8 In a metric space, the derived set of a set is a closed set.

Proof. Let (X, d) be a metric space, and $A \subseteq X$. Let x be any point in $X - D(A)$. Then x is not a limit point of A. So there exists a real number $\varepsilon > 0$ such that the open sphere $S(x, \varepsilon)$ does not contain any point of A other than x. If $S(x, \varepsilon) \cap D(A) \neq \emptyset$, choose a point y in $S(x, \varepsilon) \cap D(A)$. Since y is a limit point of A, the open sphere $S(y, \varepsilon_1)$, where $\varepsilon_1 = \varepsilon - d(x, y)$, contains infinitely many points of A. Since $S(y, \varepsilon_1) \subseteq S(x, \varepsilon)$, it means that $S(x, \varepsilon)$ contains a point of A other than x, which is a contradiction. Therefore, $S(x, \varepsilon) \subseteq X - D(A)$. Since x is arbitrary, $X - D(A)$ is an open set, and $D(A)$ is a closed set.

Definition 2.8.9 Let (X, d) be a metric space, and A be a subset of X. The closure of A, denoted by $\overline{A}$, is defined by

$$\overline{A} = A \cup D(A).$$

Examples 2.8.10

1. The closure of $\mathbb{Q}$ in the real line $\mathbb{R}$ is $\mathbb{R}$ since $D(\mathbb{Q}) = \mathbb{R}$. The closure of the $A = \{1/n : n \in \mathbb{Z}_+\}$ in $\mathbb{R}$ is the set $A \cup \{0\}$. If $A = [0, 1[$, then $\overline{A} = [0, 1] \cdot \overline{\mathbb{Z}} = \mathbb{Z}$ since no real number can be a limit point of $\mathbb{Z}$.

2. Let us consider the metric space $\mathbb{R}^2$. If $S = \{(x, y) \in \mathbb{R}^2 : y = 2x\}$, then $\overline{S} = S \cdot (\overline{\mathbb{Q} \times \mathbb{Q}}) = \mathbb{R}^2$. If $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, then $\overline{S} = S$. If $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$, then $\overline{S} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. If $S = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1$ and y is a real number$\}$, then $\overline{S} = S$.

3. In the real $\mathbb{R}$, let $S = \{1/m + 1/n : m, n \in \mathbb{Z}_+\}$. Then $D(S) = \{1/n : n \in \mathbb{Z}_+\} \cup \{0\}$.

4. In a discrete metric space, the closure of a set A is A itself.

5. In any metric space (X, d), $\overline{A} = A$ for any finite subset A of X.

6. In $\mathbb{R}^n$, the closure of an open sphere $S(x, r)$ is the closed sphere $S[x, r]$.

Theorem 2.8.11 The closure of a set is a closed set.

Proof. Let (X, d) be a metric space, and A be a subset of X. We show that $\overline{A}$ contains all its limit points. Assume that a point x in $x - \overline{A}$ is a limit point of $\overline{A}$. Let $\varepsilon > 0$ be a real number. Then the open sphere $S(x, \varepsilon/2)$ contains a point y of $\overline{A}$. Now either $y \in A$, or y is a limit point of A. In the later case, the open sphere $S(y, \varepsilon/2)$ contains a point z in A which is different from y. By the inequality $d(x, z) \leq d(x, y) + d(y, z)$, it follows that $z \in S(x, \varepsilon)$. In the first case where y is in A, the open sphere $S(x, \varepsilon)$ contains y. Thus, every open sphere centered at x intersects with A. But then x is a limit point of A, and so x is in $\overline{A}$. This is against our assumption, and the proof is complete.

Theorem 2.8.12 Let (X, d) be a metric space, and A and B subsets of X. Then the following statements hold.

1. $\overline{A}$ is the intersection of all closed supersets of A.
2. A is closed if and only if $\overline{A} = A$.
3. If $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$.
4. $(\overline{A \cup B}) = \overline{A} \cup \overline{B}$.
5. $(\overline{A \cap B}) \subseteq \overline{A} \cap \overline{B}$.
6. $\overline{\overline{A}} = \overline{A}$.

Proof. 1. By the definition of closure of a set, $A \subseteq \overline{A}$. Let B be any closed set containing A. Let x be any point in $B' = X - B$. Then x is not a limit point of B and hence of A. Thus, $\overline{A} \subseteq B$, and the assertion follows. The other parts are left to the reader.

Theorem 2.8.13 The diameter of a set is equal to the diameter of its closure.

Proof. Let (X, d) be a metric space, and A be its subset. If $A = \emptyset$, then $\overline{A} = \emptyset$, and $d(A) = d(\overline{A})$. If A is non-empty, since $d(A) \leq d(\overline{A})$, we show the inequality $d(\overline{A}) \leq d(A)$. Let x and y be any two points in $\overline{A}$. Then for every real number $\varepsilon > 0$, there exist points x' and y' in A such that $d(x, x') < \varepsilon$, and $d(y, y') < \varepsilon$. Therefore,

$$d(x, y) \leq d(x, x') + d(x', y') + d(y', y) < d(A) + 2\varepsilon.$$

Since ε is arbitrary, it follows that $d(\overline{A}) \leq d(A)$.

Theorem 2.8.14 Every infinite and bounded subset of $\mathbb{R}$ has a limit point.

Proof. Let A be an infinite and bounded set in $\mathbb{R}$. Therefore, by the order completeness property of $\mathbb{R}$, supremum of A and infimum of A exist. Let $k = \inf A$ and $k' = \sup A$. Let us define

$$S = \{x \in \mathbb{R} : x \text{ exceeds only finite number of elements of } A\}$$

Then $k \in S$, and k' is an upper bound of S. Therefore, $\sup S$ exists. Let $p = \sup S$. Let $\varepsilon > 0$ be a real number. Then $]p - \varepsilon, p + \varepsilon[$ contains infinitely many members of A. Thus, p is a limit point of A.

Corollary 2.8.15 Every infinite and bounded subset of $\mathbb{R}^n$ has a limit point in $\mathbb{R}^n$.

Proof. The proof is by induction. The base case of $n = 1$ is the preceding theorem. In the inductive step, assume that every infinite and bounded set in $\mathbb{R}^k$ has a limit point in $\mathbb{R}^k$. Let A be an infinite and bounded set in $\mathbb{R}^{k+1}$. Let

$$A_k = \{(x_1, \ldots, x_k) : (x_1, \ldots, x_k, x_{k+1}) \in A\}.$$

Case (i) A_k is finite. Then there exists a point $(y_1, \ldots, y_k) \in A_k$ such that the set

$$\{y_{k+1} : (y_1, \ldots, y_k, y_{k+1}) \in A\}$$

is infinite and bounded. So, by the preceeding theorem, it has a limit point y. Then $(y_1, \ldots, y_k, y)$ is a limit point of A.
Case (ii) A_k is infinite. Since A_k is bounded, by the induction hypothesis, A_k has a limit point $y = (y_1, \ldots, y_k)$. Suppose that there exists a neighbourhood N of y such that the set

$$B_{k+1} = \{z : (x_1, \ldots, x_k) \in N \text{ and } (x_1, \ldots, x_{k,z}) \in A\}$$

is finite. Now, if for each $z \in B_{k+1}$, there exists a neighbourhood $N_z \subseteq N$ of y such that the set

$$\{(x_1, \ldots, x_k) \in N : (x_1, \ldots, x_k, z) \in A\} \cap N_z$$

is finite, then the intersection of these is a neighbourhood of y which contains only finite number of elements of N and hence of A_k. Then y is not a limit point of A_k, a contradiction. So, there exists a point $z \in B_{k+1}$ such that for each neighbourhood N' of y, the set $\{(x_1, \ldots, x_k) \in N : (x_1, \ldots, x_k, z) \in A\} \cap N'$ is infinite. Then $(y_1, \ldots, y_k, z)$ is a limit point of A. So we assume that for each neighbourhood N of y, the set B_{k+1} is infinite. In particular, for each neighbourhood of the form $S(y, 1/n)$, $n \in \mathbb{Z}_+$, the corresponding B_{k+1} is written as B^n_{k+1}. Since B^n_{k+1} is infinite and bounded, it has a limit point by the preceeding theorem. Let $z_n = \sup\{z : z \text{ is a limit point of } B^n_{k+1}\}$. Then the sequence $< z_n >$ is non-increasing and bounded in $\mathbb{R}$ and thus converges to a point z_0. Then $(y_, \ldots, y_k, z_0)$ is a limit point of A.

EXERCISE

1. Let $A_1, A_2, \ldots$ be subsets of a metric space. Show that $\bigcup_{i=1}^{\infty} \overline{A}_i \subseteq \overline{\left(\bigcup_{i=1}^{\infty} A_i\right)}$, and give an example to show that this inclusion can be proper.

2. Given an example of a metric space in which every uncountable subset has a limit point but no countable subset has a limit point.

3. Let A and B be subsets in a metric space X such that $A \cap \overline{B} = \overline{A} \cap B = \emptyset$. Show that there exists disjoint open set G and H such that $A \subseteq G$, and $B \subseteq H$.

2.9 Adherent Points and Isolated Points

Definition 2.9.1 Let (X, d) be a metric space, and A be a subset of X. A point x in X is said to be an adherent point of A if every neighbourhood of x intersects with A. The set of all adherent points of a set A is denoted by $\mathrm{Adh}(A)$.

Examples 2.9.2

1. The set of all adherent points of the sets $]a, b[$, $\mathbb{Z}, Q, \{1/n : n \in \mathbb{Z}_+\}$ in $\mathbb{R}$ are the sets $[a, b], \mathbb{Z}, \mathbb{R}$ and $\{1/n : n \in \mathbb{Z}_+\} \cup \{0\}$, respectively.

2. In a discrete metric space (X, d), the set of adherent points of a set A coincides with A itself.

3. In a metric space (X, d), the set of adherent points of a set A contains A.

4. In $\mathbb{R}^2$, a point $x_0 \in \mathbb{R}^2$ is an adherent point of the line $y = mx$, where m is a real number, if and only if the point lies on the line $y = mx$.

The following theorems are immediate consequences of the above definition.

Theorem 2.9.3 Let (X, d) be a metric space, and A be a subset of X. A point x in X is an adherent point of A if and only if for each real number $\varepsilon > 0$, the open sphere $S(x, \varepsilon)$ intersects with A.

Theorem 2.9.4 Let (X, d) be a metric space, and A be a subset of X. If a point x in X is a limit point of A, then x is an adherent point of A.

The converse of the above theorem is not true. In the real line $\mathbb{R}$, all the members of the set $A = \{1/n : n \in \mathbb{Z}_+\}$ are adherent points of A but none of the members is a limit point of A.

Theorem 2.9.5 Let (X, d) be a metric space, and A be a subset of X. A point $x \in X$ is an adherent point of A if and only if $x \in \overline{A}$, the closure of A, that is $\text{Adh}(A) = \overline{A}$.

Proof. If $x \in \overline{A}$, then either $x \in A$, or x is a limit point of A. In either case x is an adherent point of A. Conversely, if x is an adherent point of A, then if x does not belong to A, then every neighbourhood of x contains a point of A other than x; so x is a limit point of A. Thus, $x \in \overline{A}$.

Theorem 2.9.6 Let (X, d) be a metric space, and A be a subset of X. Then $x \in \text{Adh}A$ if and only if $d(x, A) = 0$.

Proof. A point $x \in \text{Adh}(A)$ if and only if for each real $\varepsilon > 0$, there exists a point y in A such that $y \in S(x, \varepsilon)$, that is, $d(x, y) < \varepsilon$ if and only if $\inf\{d(x, y) : y \in A\} < \varepsilon$ if and only if $d(x, A) = 0$, since ε is arbitrary.

Theorem 2.9.7 Let (X, d) be a metric space, and A be a subset of X. Then $\text{Adh}(A) = \bigcap\{D_r(A) : r > 0\}$, where $D_r(A) = \{x \in X : d(x, A) < r\}$ (see Example 2.4.2(8)).

Proof. Let x be any point in $\text{Adh}(A)$. By Theorem 2.9.6, $d(x, A) = 0$. Therefore, x is in every $D_r(A)$. Conversely, if $x \in \bigcap\{D_r(A) : r > 0\}$, then $x \in D_r(A)$ for every real number $r > 0$. Therefore, $d(x, A) < r$ for every real number $r > 0$. Thus, $d(x, A) = 0$, and $x \in \text{Adh}(A)$.

Theorem 2.9.8 In a metric space, any closed set A is the intersection of a decreasing sequence of open sets; any open set A is the union of an increasing sequence of closed sets.

Proof. Let A be a closed set. By Theorem 2.5.11, the sequence $< D_{1/n}(A) >$ is a sequence of open sets which is decreasing, that is, $D_{1/m+1}(A) \subseteq D_{1/m}(A)$ for all $m \in \mathbb{N}$. Also $A \subseteq D_{1/n}(A)$ for all $n \in \mathbb{Z}_+$ so that $A \subseteq \bigcap_{n=1}^{\infty} D_{1/n}(A)$. Conversely, if x is any point in $\bigcap_{n=1}^{\infty} D_{1/n}(A)$, then $d(x, A) = 0$. Therefore, $x \in \text{Adh}(A) = \overline{A} = A$ since A is a closed set. We leave it to the reader to prove the second part of the theorem.

Definition 2.9.9 Let (X, d) be a metric space, and $A \subseteq X$. A point x in A is said to be an isolated point of A if there exists a neighbourhood of x which does not contain any point of A other than x.

An isolated point of a set A is an adherent point but not conversely. Let us consider the metric space $\mathbb{R}$, and $A =]0, 1[$. Then 0.5 is an adherent point of A but it is not an isolated point in A.

Examples 2.9.10

1. Every point of a finite set in any metric space is an isolated point of that set.
2. In a discrete metric space, every point of any subset is an isolated point.
3. In the real line $\mathbb{R}$, if $A =]0, 1[\cup \{2\}$, then 2 is an isolated point in A.

2.10 Interior, Boundary and Exterior of a Set

In the preceding section, we formed the closure of a set A by including in A its limit points. Similarly, the interior of a set, boundary of a set and exterior of a set are formed in the following manner.

Definition 2.10.1 Let (X, d) be a metric space, and A be a subset of X. A point x in X is said to be an interior point of A if A is a neighbourhood of x. The interior of A, denoted by A^0 or $\operatorname{Int} A$, is the set of all interior points of A.

Examples 2.10.2

1. In the real line $\mathbb{R}$, the interior of any open interval $]a, b[$ is $]a, b[$ itself. The interior of $\mathbb{Q}$ in $\mathbb{R}$ is empty. The interior of any finite subset of $\mathbb{R}$ is empty. The interior of $S = \{1/n : n \in \mathbb{Z}_+\}$ is empty. The interior of the closed interval $[a, b], a < b$, is the open interval $]a, b[$.

2. Let us consider the metric space $\mathbb{R}^2$. If $S = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, \text{ and } 0 < y < 1\}$ then $S^0 = S$. If $S_1 = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, \text{ and } 0 < y < 1\}$, then $S_1^0 = S$. Also $\mathbb{Q} \times \mathbb{Q}^0$ is empty. If $S_2 = \{(x, y) \in \mathbb{R}^2 : y = mx, \text{ where } m \text{ is a real number}\}$, then S_2^0 is empty.

3. The interior of any subset A of a discrete metric space is A itself.

4. The interior of the following subsets of the complex plane $\mathbb{C} : \{z \in \mathbb{C} : |z| < 1\}$, $\{z \in \mathbb{C} : |z| \leq 1\}$, and $\{z \in \mathbb{C} : \operatorname{Re} z \text{ is rational}\}$ are the sets $\{z \in \mathbb{C} : |z| < 1\}$, $\{z \in \mathbb{C} : |z| < 1\}$ and the empty set, respectively.

Theorem 2.10.3 Let (X, d) be a metric space, and A and B be subsets of X. Then the following statements hold.

1. A^0 is an open set.
2. A^0 is the largest open set contained in A; A^0 is the union of all open sets contained in A.
3. A is open if and only if $A^0 = A$.
4. $A \subseteq B$ implies that $A^0 \subseteq B^0$.
5. $A^0 \cup B^0 \subseteq (A \cup B)^0$.
6. $(A \cap B)^0 = A^0 \cap B^0$.
7. $(A^0)^0 = A^0$.

Proof. We prove here only the statements (1) and (2), and leave the rest to the reader.

1. Let x be a point in A^0. Then there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq A$. We show that $S(x, \varepsilon) \subseteq A^0$. Let y be any point in $S(x, \varepsilon)$. Let $\varepsilon_1 = \varepsilon - d(x, y)$. Then the open sphere $S(y, \varepsilon_1) \subseteq S(x, \varepsilon)$. So y is an interior point of A. Since y is an arbitrary point of $S(x, \varepsilon)$, $S(x, \varepsilon)$ is contained in A^0. Therefore, A^0 is an open set.

2. Let B be a subset of A, and B be open in X. Let x be any point in B. Then there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq B$. Then $S(x, \varepsilon) \subseteq A$. So x is an interior point of A. This means that $B \subseteq A^0$. Since A^0 is an open set, it is the largest open set contained in A.

In the statement (5) of the above theorem, the containment can not be replaced by equality. For instance, in $\mathbb{R}^2$, let $A = \{(x, y) =\in \mathbb{R}^2 : 0 \leq x \leq 1$, and y is arbitrary$\}$, and $B = \{(x, y) \in \mathbb{R}^2 : 1 \leq x \leq 2$, and y is arbitrary$\}$. Then

$$(A \cup B)^0 = \{(x, y) \in \mathbb{R}^2 : 0 < x < 2, \text{ and } y \text{ is arbitrary}\}$$

while

$$A^0 \cup B^0 = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, \text{ or } 1 < x < 2, \text{ and } y \text{ is arbitrary}\}$$

which is not equal to $(A \cup B)^0$.

Theorem 2.10.4 Let (X, d) be a metric space, and A be a subset of X. Then

$$\text{(i) } (A^0)' = \overline{(A')} \quad \text{(ii) } \overline{A} = ((A')^0)' \quad \text{(iii) } (\overline{A})' = (A')^0$$

where dashes denote the complements in X.

Proof. (i) The point x is in $(A^0)'$ if and only if x is not in A^0 if and only if for each real number $\varepsilon > 0$, the open sphere $S(x, \varepsilon)$ is not contained in A if and only if for each real number $\varepsilon > 0$, the open sphere $S(x, \varepsilon)$ intersects with A' if and only if either $x \in A'$ or x is a limit point of A' if and only if $x \in \overline{(A')}$.

(ii) In the statement (i), replace A by A'.

(iii) Take the complement in the equality of statement (ii).

Definition 2.10.5 Let (X, d) be a metric space, and A be a subset of X. An interior point of $X - A$ is called an exterior point of A, and interior of $X - A$ is called the exterior of A, denoted by $\text{Ext}(A)$.

Theorem 2.10.6 Let (X, d) be a metric space, and A be a subset of X. A point $x \in X$ is an exterior point of A if and only if $d(x, A) > 0$.

Proof. If $d(x, A) > 0$, then the open sphere $S(x, d(x, A)) \subseteq X - A$. Thus, x is an interior point of $X - A$, and hence $x \in \text{Ext}(A)$. Conversely, if x is an exterior point of A, then there exists a real number ε such that $S(x, \varepsilon) \subseteq X - A$. This means that $d(x, y) \geq \varepsilon$ for all y in A, that is, $d(x, A) > 0$.

Theorem 2.10.7 Let (X, d) be a metric space, and A be a subset of X. Then $\text{Ext}(A) = (\overline{A})'$.

Proof. The proof is left to the reader.

Definition 2.10.8 Let (X, d) be a metric space, and A be a subset of X. A point x in X is said to be a boundary point (or frontier point) of A if every open sphere centered at x intersects both A and the complement of A in X. The set of boundary points of A is denoted by $Bd(A)$.

It follows from the above definition that a boundary point of A is neither an interior nor an exterior point of A.

Examples 2.10.9

1. Let us consider the real line $\mathbb{R}$. If $I =]0, 1[$, then $Bd(I) = \{0, 1\}$. Also $Bd(\mathbb{Q}) = \mathbb{R}$, $Bd(\mathbb{R} - \mathbb{Q}) = \mathbb{R}$, $Bd(\emptyset) = \emptyset$, $Bd(\mathbb{N}_0) = \mathbb{N}_0$.

2. In $\mathbb{R}^2$, if

$$A = \{(x, y) \in \mathbb{R}^2 : x > 0\},$$
$$B = \{(x, y) \in \mathbb{R}^2 : y = mx, \text{ where } m \text{ is a real number}\},$$
$$C = \text{Any finite subset of } \mathbb{R}^2,$$
$$D = \text{A unit circle},$$

then $Bd(A) = \{(x, y) \in \mathbb{R}^2 : x = 0\}$, $Bd(B) = B$, $Bd(C) = C$ and $Bd(D) = D$.

The following theorem summarizes elementary facts about the boundary of a set, and the proofs are left to the reader.

Theorem 2.10.10 Let (X, d) be a metric space, and A be a subset of X. Then

1. $Bd(A) = Bd(X - A)$.
2. $Bd(A) = \overline{A} \cap \overline{(X - A)}$.
3. $Bd(A) = \overline{A} - A^0$.
4. $Bd(A) = \overline{(x - A)} - (X - A)^0$.
5. $X - Bd(A) = A^0 \cup (X - A)^0$.
6. $\overline{A} = A^0 \cup Bd(A) = A \cup Bd(A)$.
7. $A^0 \cap Bd(A) = \emptyset$.
8. $A^0 = A - Bd(A)$.
9. A is closed if and only if $Bd(A) \subseteq A$.
10. A is open if and only if $A \cap Bd(A) = \emptyset$.
11. $Bd(A) = \emptyset$ if and only if A is both open and closed.

2.11 Dense Sets, Perfect Sets and Separability

Definition 2.11.1 Let (X, d) be a metric space. A subset A of X is said to be dense in X if the closure of A coincides with X, that is, every point of X is an adherent point of A.

Examples 2.11.2

1. The set of rational numbers is dense in $\mathbb{R}$. Also the set of irrational numbers is dense in $\mathbb{R}$.
2. The sets $\mathbb{Q}^n$ and $\mathbb{R}^n - \mathbb{Q}^n$ are both dense in $\mathbb{R}^n$.
3. In a metric space (X, d), the underlying set X is dense in X.
4. A subset A of a discrete metric space (X, d) is dense in X if and only if $A = X$.

The following theorem gives some equivalent conditions for a dense set.

Theorem 2.11.3 Let (X, d) be a metric space, and A be a subset of X. The following statements are equivalent.

1. A is dense in X.

2. The only closed superset of A is X.
3. A intersects every non-empty open set in X.

Proof. (1) implies (2). Suppose that A is dense in X. Then $\overline{A} = X$. Since $\overline{A}$ is the smallest closed superset of A, the conclusion (2) follows.

(2) implies (3). Let G be a non-empty open set in X which does not intersect with A that is, $G \subseteq A' = X - A$. Let x be any point in G. Since G is open, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G$. So that $S(x, \varepsilon) \subseteq A'$. Therefore, $A \subseteq (S(x, \varepsilon))'$, the complement of $S(x, \varepsilon)$ in X. Now $(S(x, \varepsilon))'$ is a closed set containing A, and not equal to X. This is against our hypothesis.

(3) implies (1). If A intersects every non-empty open set in X, in particular, A intersects with every open sphere centered at a point in $A' = X - A$. Thus, every point of A' is a limit point of A, and hence $\overline{A} = X$.

Definition 2.11.4 A subset A of a metric space (X, d) is said to be dense in itself if every point of A is a limit point of A, that is, $A \subseteq D(A)$. A subset A of a metric space (X, d) is said to be perfect if A is both closed and dense in itself, that is, $A = D(A)$.

Examples 2.11.5

1. The closed interval $[0, 1]$ in $\mathbb{R}$ is a perfect set.
2. In a discrete metric space, no subset is a perfect set.

Intuitively, a set is dense in a metric space X if its elements are arbitrarily close to each element of X. It is also useful to consider sets whose elements are sparsely spread in X.

Definition 2.11.6 A subset A of a metric space (X, d) is said to be nowhere dense in X if the interior of the closure of A is empty, that is, $(\overline{A})^0 = \emptyset$.

Examples 2.11.7

1. Any finite set in $\mathbb{R}$ is a nowhere dense set.
2. The sets $\mathbb{N}_0$, $\mathbb{Z}$, $A = \{1/n : n \in \mathbb{Z}_+\}$ are nowhere dense sets in $\mathbb{R}$.
3. Any line in $\mathbb{R}^2$ is a nowhere dense set in $\mathbb{R}^2$. Any circle, hyperbola, parabola or ellipse in $\mathbb{R}^2$ is a nowhere dense set.
4. The empty set is the only nowhere dense set in a discrete metric space.

The following theorem gives some equivalent conditions for a nowhere dense set.

Theorem 2.11.8 Let A be a subset of a metric space (X, d). The following statements are equivalent.

1. A is nowhere dense in X.
2. $\overline{A}$ does not contain any non-empty open set in X.
3. For each non-empty open set G in X, there is a non-empty open set G^* in X such that $G^* \subseteq G$, and $G^* \cap \overline{A} = \emptyset$.
4. For each non-empty open set G in X, there is a a non-empty open set G^* in X such that $G^* \subseteq G$, and $G^* \cap A = \emptyset$.

Moreover, if A is nowhere dense in X, then for each open set G, there exists an open sphere S such that $\overline{S} \subseteq G$ and $\overline{S} \cap A = \emptyset$.

Proof. (1) implies (2). If G is any non-empty open set contained in $\overline{A}$, then every point of G is an interior point of $\overline{A}$, and so $(\overline{A})^0 \neq \emptyset$.

(2) implies (3). Let G be any non-empty open set in X. Then, by our assumption, $G \cap (\overline{A})' \neq \emptyset$. Also $G \cap (\overline{A})'$ is an open set. Then for each point x in $G \cap (\overline{A})'$, there is a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G \cap (\overline{A})'$. Let us take $G^* = S(x, \varepsilon)$.

(3) implies (4) Obvious.

(4) implies (1). Suppose, if possible, $(\overline{A})^0 \neq \emptyset$. Let $G = (\overline{A})^0$. Since G is open, by our assumption, there exists a non-empty open set G^* in X such that $G^* \subseteq G$, and $G^* \cap A = \emptyset$. Let x be a point in G^*. Then there is a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G^*$. Thus, $S(x, \varepsilon) \cap A = \emptyset$. This means that x is neither a point of A nor a limit point of A, that is, x is not in $\overline{A}$. This is a contradiction since $x \in G^* \subseteq G \subseteq \overline{A}$.

The following theorem relates the notions of a dense set and of a nowhere dense set.

Theorem 2.11.9 Let (X, d) be a metric space. A subset A of X is nowhere dense in X if and only if $X - \overline{A}$ is dense in X.

Proof. Let x be a point in $\overline{A}$. Since $(\overline{A})^0 = \emptyset$, x is not an interior point of $\overline{A}$. That means for each real number $\varepsilon > 0$, $S(x, \varepsilon) \cap (X - \overline{A}) \neq \emptyset$. Therefore, x is a limit point of $X - \overline{A}$, that is, $\overline{(X - \overline{A})} = X$. Conversely, assume that $X - \overline{A}$ is dense in X. Then every point of $\overline{A}$ is a limit point of $X - \overline{A}$. That is, for no real $\varepsilon > 0$, the open sphere $S(x, \varepsilon)$ is contained in $\overline{A}$. Therefore, $(\overline{A})^0 = \emptyset$, and A is nowhere dense in X.

Corollary 2.11.10 Let (X, d) be a metric space, and A be a closed subset of X. Then A is nowhere dense in X if and only if $X - A$ is dense in X.

Corollary 2.11.11 Let (X, d) be a metric space, and A be an open set in X. Then A is dense in X if and only if $X - A$ is nowhere dense in X.

Proof. Replace in Corollary 2.11.10, A by $X - A$.

Theorem 2.11.12 A finite union of nowhere dense sets in a metric space is a nowhere dense set.

Proof. Let $\{A_i\}$ be a finite family of nowhere dense sets in a metric space (X, d), and A be the union of this family. If the family $\{A_i\}$ is empty, or contains only the empty set $\emptyset$, then A is empty, and it is a nowhere dense set. So let $\{A_i\}$ be a non-empty family containing non-empty nowhere dense sets $A_1, A_2, \ldots, A_n$ in X. Let G be any non-empty open set in X. By Theorem 2.11.8, there exists a non-empty open set G_1^* in X such that $G_1^* \subseteq G$, and $G_1^* \cap A_1 = \emptyset$. Again, by Theorem 2.11.8, there exists a non-empty open set G_2^* such that $G_2^* \subseteq G_1^*$, and $G_2^* \cap A_2 = \emptyset$. Continuing in this way, we get subsets $G_1^*, G_2^*, \ldots, G_n^*$ such that

$$G_n^* \subseteq G_{n-1}^* \subseteq \ldots G_1^* \subseteq G, \text{ and } G_i^* \cap A_i = \emptyset \text{ for each } i = 1, 2, \ldots, n.$$

Now it follows that $G_n^* \subseteq G$, and $G_n^* \cap \left(\bigcup_{i=1}^{n} A_i \right) = \emptyset$.

We can not extend the result of the above theorem to arbitrary unions not even to countable unions of nowhere dense sets. For example, consider the real line $\mathbb{R}$, and the sets

$$F_n = \{p/n : p \in \mathbb{Z}\} \text{ for all } n \in N.$$

Then $\bigcup_{n=1}^{\infty} F_n = \mathbb{Q}$. F_n's are nowhere dense sets, and $\mathbb{Q}$, of course, is not a nowhere dense set in $\mathbb{R}$.

Definition 2.11.13 A metric space (X, d) is said to be separable if it has a subset A which is countable and dense in X.

Examples 2.11.14

1. In the real line $\mathbb{R}$, the set of rational numbers $\mathbb{Q}$ is countable and dense in $\mathbb{R}$. Therefore, $\mathbb{R}$ is separable.

2. The set $\mathbb{Q}^n$ is countable and dense in $\mathbb{R}^n$. Thus, $\mathbb{R}^n$ is separable.

3. Any finite metric space is separable. Any countable metric space is separable.

4. A discrete metric space (X, d) is separable if and only if X is countable.

5. The l_p spaces (Example 2.1.20) are separable. Let us define a subset of l_p:

$$Y = \{\langle x_1, x_2, \ldots, x_n, 0, 0, \ldots\rangle : x_i \in \mathbb{Q}, 1 \leq i \leq n, \text{ and } n \in \mathbb{Z}_+\}.$$

Then $Y = \bigcup_{r=1}^{\infty} Y_r$, where, for $r \in \mathbb{Z}_+$,

$$Y_r = \{\langle x_1, x_2, \ldots, x_r, 0, 0, \ldots\rangle : x_i \in Q, \text{ for } 1 \leq i \leq r\}$$

Each Y_r is countable. For $(x_1, x_2 \ldots, x_n) \in \mathbb{Q}^n$, the mapping $(x_1, x_2, \ldots, x_n) \to \langle x_1, x_2, \ldots, x_n, 0, 0, \ldots\rangle$ is a one-to-one correspondence between $\mathbb{Q}^n$ and Y_n, and $\mathbb{Q}^n$ is a countable set. Thus, Y, being a countable union of countable sets, is a countable set. Now we show that Y is dense in l_p. Let $x = \langle x_n\rangle$ be a point in l_p. Let $\varepsilon > 0$ be a real number. Then there exists $r \in \mathbb{Z}_+$ such that

$$\sum_{i=r+1}^{\infty} |x_i|^p < \frac{\varepsilon^p}{2}.$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists $y_1, y_2, \ldots, y_r$ in $\mathbb{Q}$ such that

$$|x_i - y_i| < \frac{\varepsilon}{(2r)^{1/p}} \text{ for all } i = 1, 2, \ldots, r.$$

Then the sequence $y = \langle y_1, y_2, \ldots, y_r, 0, 0, \ldots\rangle$ is in Y, and

$$d_p(x, y) = \left[\sum_{i=1}^{r} |x_i - y_i|^p + \sum_{i=r+1}^{\infty} |x_i|^p\right]^{1/p} < \varepsilon.$$

Therefore, Y is dense in l_p, and l_p is separable. The proof of the complex case is left to the reader (In the definition of Y, let x_i be the complex numbers with rational co-ordinates).

Example 2.11.15 The l_∞-space in Example 2.1.23 is not separable. To show this, let $x = \langle x_i\rangle$ be a sequence such that $x_i = 0$ or 1 for all i. Then $x \in l_\infty$. With x we associate a real number $\hat{x}$ such that the binary expansion of $\hat{x}$ is $.x_1x_2x_3\ldots$. Then $\hat{x} \in [0, 1]$. Since

each real number in $[0,1]$ has a unique binary expansion, distinct real numbers in $[0,1]$ give rise to distinct sequences of zeros and ones. Since $[0,1]$ is uncountable, the set of sequences consisting of zeros and ones in l_∞ is uncountable. Also, the metric of l_∞ shows that the distance between two such distinct sequences is 1. If, we let each of these sequences to be the centre of an open sphere of radius $1/3$, then there will be uncountable non-intersecting open spheres. If M is any dense set in l_∞, then it will be intersecting with each of these open spheres. Therefore, M is uncountable. Since M is an arbitrary set, l_∞ can not be separable.

EXERCISE

1. Show that a non-empty perfect set in $\mathbb{R}$ is uncountable.
2. Show that the metric space in Example 2.1.24 is separable.
3. Let A_1 and A_2 be open and dense sets in a metric space X. Then $A_1 \cup A_2$ and $A_1 \cap A_2$ are open and dense in X.
4. A set A is dense in a metric space X if and only if for each point x in X, there exists an infinite sequence $< x_n >$ in A converging to x in X.
5. Let (X,d) be a separable metric space. Prove that the set of all open sets of (X,d) has cardinality not greater than the cardinality of $\mathbb{R}$. Deduce that X also has cardinality not greater than that of $\mathbb{R}$.
6. Let (X,d) be a separable metric space, and $\{A_\lambda\}_{\lambda\in\Lambda}$ be a family of open sets in X such that $A_\lambda \cap A_{\lambda'} = \emptyset$ for $\lambda \neq \lambda'$, and $\bigcup\{A_\lambda : \lambda \in \Lambda\} = X$. Prove that Λ is a countable set. Let $Y \subseteq X$ such that every point of Y is an isolated point. Show that Y is countable. Also deduce that any uncountable subset Z of X must contain a limit point of Z.
7. Let G be an open set in a metric space X, and A be a dense subset of X. Show that $G \subseteq (\overline{G \cap A})$.
8. Let X be the set of all real or complex sequences $< a_n >$ such that the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, and

$$d(a,b) = \sum_{n=1}^{\infty} |a_n - b_n|$$

for all $a =< a_n >$ and $b =< b_n >$ in X. Show that (X,d) is a metric space. Also, (X,d) is separable.

9. In a metric space (X,d), a set A is said to be ε-rare for a real number $\varepsilon > 0$ if $x, y \in A$, $x \neq y$ implies that $d(x,y) \geq \varepsilon$. Show
   (a) if for each real number $\varepsilon > 0$, each ε-rare set is countable, then the metric space is separable.
   (b) if for some real number $\varepsilon > 0$, there exists an uncountable ε-rare set, then the metric space is not separable.

10. Is there a non-empty perfect set in $\mathbb{R}$ which contains no rational number?

11. Show that if G is an open set in a metric space X, then $\overline{G} - G$ is nowhere dense.

2.12 Second Axiom of Countability

In the theory of metric spaces, a particular collection of open sets has played a special role, viz., the collection of open spheres. For instance, sometimes it happens that if some statement is true for open spheres, then it is true in general for all open sets. In various other settings, other collections of open sets play the important special roles. In this context, the following concept is of interest.

Definition 2.12.1 A collection $\mathbb{B}$ of open subsets of a metric space (X, d) is called an open base for X if every open set in X is the union of some subcollection of $\mathbb{B}$.

Since every open set in a metric space (X, d) is the union of a collection of open spheres, the collection of all open spheres in (X, d) is an open base for X.

Theorem 2.12.2 Let (X, d) be a metric space, and $\mathbb{B}$ be a collection of open sets in X. Then $\mathbb{B}$ is an open base for X if for every open set U in X and every point x in U, there exists a set V in $\mathbb{B}$ such that $x \in V \subseteq U$.

Proof. For each open set U in X, let $\mathbb{B}_U$ be the subcollection $\{V \in \mathbb{B} : V \subseteq U\}$. If $\mathbb{B}$ is an open base for X, then, by the definition of an open base, U is the union of sets in $\mathbb{B}_U$. Therefore, if $x \in U$, then there exists a set V in $\mathbb{B}_U$ such that $x \in V \subseteq U$. Conversely, assume that for a given open set U in X and a point x in U, there exists an open set V_x in $\mathbb{B}$ such that $x \in V_x \subseteq U$. We take the collection of all V_x's existing in this way for each x in U, and denote it by $\mathbb{B}_U$. Let $\mathbb{B} = \bigcup\{\mathbb{B}_U : U \text{ is an open set in } X\}$. Then $\mathbb{B}$ is an open base for X.

It is clear from the above definition of an open base for a metric space (X, d) that if $\mathbb{B}$ is an open base for a metric space (X, d), and $\mathcal{C}$ is another collection of open sets in X, then $\mathbb{B} \cup \mathbb{C}$ is also an open base for X. Therefore, it is sometimes of interest to seek an open base containing minimum number of open sets. The following definition is an attempt in this direction.

Definition 2.12.3 A metric space (X, d) is said to satisfy the second axiom of countability (or is a second countable space) if there exists a countable open base for X.

At this point, a reader curious to know what became of the "first" axiom of countability is counselled to be patient. This so called first axiom of countability is an important concept in the class of topological spaces which contains the class of metric spaces as a subclass. The concept of first axiom of countability does not classify metric spaces in the sense that every metric space satisfies it. We shall discuss the first axiom of countability later in this section.

The following theorem shows that the second axiom of countability is equivalent to the notion of separability in metric spaces.

Theorem 2.12.4 A metric space (X, d) satisfies the second axiom of countability if and only if it is separable.

Proof. Let (X, d) be a metric space satisfying the second axiom of countability. Let $\mathbb{B}_0$ be a countable open base for X. If the empty set is in $\mathbb{B}_0$, delete it. Now, we choose an

element x_V from each set V in $\mathbb{B}_0$. Then the set of these selected points $A = \{x_V\}_{V \in \mathbb{B}_0}$ is a countable set. We show that A is dense in X. Let y be any point in $X - A$. For any open sphere $S(y, r)$ centered at y and with radius r, there exists an open set V in $\mathbb{B}_0$ such that

$$x_V \in V \subseteq S(y, r).$$

This means that y is a limit point of A, and A is dense in X. Conversely, let (X, d) be a separable metric space. Let A be a countable and dense set in X. Let E be a countable set of positive real numbers such that $\inf E = 0$. We show that the countable collection of open spheres $\mathbb{B} = \{S(x, r) : x \in A, \text{ and } r \in E\}$ is an open base for X. Let U be an open set in X, and x be any point in U. Then there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq U$. Now there is a real number r in E such that $r < \varepsilon/2$. Such a number exists by the definition of E. Since A is dense in X, there exists an element y in A such that $d(x, y) < r$. Then the open sphere $S(y, r)$ is in $\mathbb{B}$, and x is in $S(y, r)$. Finally, we show that $S(y, r)$ is contained in U. Let z be any point in $S(y, r)$. Then

$$d(x, z) \leq d(x, y) + d(y, z) < r + r = 2r < \varepsilon.$$

So that we have

$$x \in S(y, r) \subseteq S(x, \varepsilon) \subseteq U.$$

This completes the proof. (Note that we have used the axiom of choice in this theorem.)

Examples 2.12.5

1. Since $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $\mathbb{C}^n$ are separable metric spaces, these are second countable spaces. For instance, for $\mathbb{R}$ the collection of all open intervals $]a, b[$, where a and b are rational numbers, is a countable open base.

2. Let (X, d) be a discrete metric space such that X is uncountable. Then (X, d) is not a second countable space since it is not a separable metric space.

Now we give, as promised the definition of first axiom of countability.

Definition 2.12.6 A metric space (X, d) is said to satisfy the first axiom of countability (or is a first countable space) if for each $x \in X$, there exists a countable family $\{G_i\}$ of open sets such that x is in G_i for all i, and for each open set G in X containing x there exists a set G_i in the family $\{G_i\}$ contained in G.

Theorem 2.12.7 Every metric space satisfies the first axiom of countability.

Proof. Let (X, d) be a metric space. For each $x \in X$, the family of open spheres $\{S(x, 1/n) : n \in \mathbb{Z}_+\}$ is a countable family. If G is an open set in X containing x, there is a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq G$. Now taking $n \in \mathbb{Z}_+$ such that $1/n < \varepsilon$, we get $S(x, 1/n) \subseteq G$.

EXERCISE

1. Let A be a subset of a second countable space. Show that the subset of A consisting of points which are not limit points of A is countable.

2. Let (X, d) be a metric space. A point $p \in X$ is said to be a condensation point of a set $A \subseteq X$, if every neighbourhood of p contains uncountably many points of A.

Let A be a uncountable set in $\mathbb{R}^n$, and P be the set of all condensation points of A. Show that P is perfect, and atmost countably many points of A are not in P. (Hint. Let $\{U_n\}$ be a countable base for $\mathbb{R}^n$, and $W = \bigcup\{U_n : U_n \cap A \text{ is at most countable}\}$. Show that $P = X - W$.)

3. Prove that every closed set in a separable metric space is the union of a (possibly empty) perfect set and a set which is atmost countable.

4. Give examples of proper open dense subsets in $\mathbb{R}, \mathbb{R}^n, \mathbb{C}^n$.

2.13 Subspaces

Given a metric space (X, d), the following theorem provides the basis for generating new metric spaces from (X, d).

Theorem 2.13.1 Let (X, d) be a metric space, and Y be a non-empty subset of X. Then the mapping $d_Y : Y \times Y \to \mathbb{R}$ defined by $d_Y(x, y) = d(x, y)$ for all $x, y \in Y$ is a metric on Y.

Proof. Since $d(x, y) \geq 0$, $d_Y(x, y) \geq 0$ for all x, y in Y. Also $d_Y(x, y) = 0$ if and only if $d(x, y) = 0$ if and only if $x = y$. The symmetry is obvious. To prove the triangle inequality let x, y and z be any point in Y. Then

$$\begin{aligned} d_Y(x, y) &= d(x, y) \leq d(x, z) + d(z, y) \\ &= d_Y(x, z) + d_Y(z, y) \text{ since } x, y, z \text{ are the points in } Y. \end{aligned}$$

Therefore, d_Y is a metric on Y, and (Y, d_Y) is a metric space.

In view of the above theorem, the following definition makes sense.

Definition 2.13.2 Let (X, d) be a metric space, and Y be a non-empty subset of X. Let d_Y be the restriction of d to $Y \times Y$. Then the metric space (Y, d_Y) is called a subspace of (X, d). The metric d_Y is usually denoted by d itself whenever there is no confusion. Sometimes, we just say that Y is a subspace of X when the underlying metric is understood.

Examples 2.13.3

1. The intervals $[a, b]$, $]a, b[$ and the set $\mathbb{Q}$, $\mathbb{Z}$ are subspaces of the real line $\mathbb{R}$. Whenever we refer an interval $[a, b]$, $]a, b]$, $[a, b[$, $]a, b[$ without mentioning its metric, it is the subspace metric of $\mathbb{R}$, that is, $d(x, y) = |x - y|$ for all x, y in the given interval.

2. A subspace is a metric space in its own right. Therefore, we can introduce in it all the concepts that are available in a metric space: open spheres, neighbourhoods, open sets, closed sets, limit points, etc. For instance, the interval $[0, 1]$ is a subspace of $\mathbb{R}$. The open sphere centered at 0 and with radius $1/2$ in $[0, 1]$, is the interval $[0, 1/2[$, a closed open interval. The open spheres $S(1, 1/2)$ is the interval $]1/2, 1]$, and the open sphere $S(1/2, 1/2)$ is the open interval $]0, 1[$. Recall that the open spheres in the real line $\mathbb{R}$ are open intervals. If we consider the subspace $]0, 1[$ of $\mathbb{R}$, then the closed sphere $S[1/2, 1/2]$ in $]0, 1[$ is the open interval $]0, 1[$ and the closed sphere $S[1/2, 1/4]$ is the closed interval $[1/4, 3/4]$.

Theorem 2.13.4 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). Then for each $y \in Y$ and a real number $\varepsilon > 0$, the open sphere $S(y, \varepsilon, d_Y)$ is the intersection with Y of the open sphere $S(y, \varepsilon, d)$, that is, $S(y, \varepsilon, d_Y) = S(y, \varepsilon, d) \cap Y$.

Proof. A point $x \in S(y, \varepsilon, d_Y)$ if and only if $x \in Y$ and $d_Y(y, x) < \varepsilon$ if and only if $d(y, x) < \varepsilon$ and $x \in y$ if and only if $x \in S(y, \varepsilon, d) \cap Y$.

If may be noted that if y is any point in $X - Y$, then the set $S(y, \varepsilon, d) \cap Y$ may not be an open sphere in Y. For example, consider the subspace $Y = \{(x, 0) \in \mathbb{R}^2 : x \text{ is a real number}\}$ of $\mathbb{R}^2$. Then $S((0, 1), 1) \cap Y$ is the empty set.

Theorem 2.13.5 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of X. A set G is open in Y if and only if there exists an open set G^* in X such that $G = G^* \cap Y$.

Proof. Let G be an open set in Y. For each $x \in G$, there exists a real number $\varepsilon_x > 0$ such $S(x, \varepsilon_x, d_Y) \subseteq G$. By Theorem 2.13.4, $S(x, \varepsilon_x, d_Y) = S(x, \varepsilon_x, d) \cap Y$. Let G^* be the union of open spheres $S(x, \varepsilon_x, d)$ obtained as above. Then G^* is an open set in X, and

$$G^* \cap Y = \left(\bigcup_{x \in G} S(x, \varepsilon_x, d) \right) \cap Y = \bigcup_{x \in G} (S(x, \varepsilon_x, d)) \cap Y) = \bigcup_{x \in G} S(x, \varepsilon_x, d_Y) = G.$$

Conversely, by Theorem 2.13.4, for each $y \in Y$ and a real number $\varepsilon > 0$, the set $S(y, \varepsilon, d) \cap Y$ is an open sphere in Y. If G^* is an open set in X, then for each $y \in G^* \cap Y$, there exists a real number $\varepsilon > 0$ such that $S(y, \varepsilon, d) \subseteq G^*$. Therefore, $S(y, \varepsilon, d) \cap Y \subseteq G^* \cap Y$, and so $G^* \cap Y$ is an open set in Y.

Corollary 2.13.6 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). Every open set G in Y is open in X if and only if Y is open in X.

Proof. The necessity part follows if we take $G = Y$. Conversely, let Y be an open set in X. Let G be an open set in Y. By the preceding theorem, there exists an open set G^* in X such that $G = G^* \cap Y$. Since $G^* \cap Y$ is open in X, G is an open set in X.

Theorem 2.13.7 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). A subset N of Y is a neighbourhood in Y of a point x in Y if and only if there exists a neighbourhood N^* in X of x such that $N = N^* \cap Y$.

Proof. Let N be a neighbourhood in Y of a point x in Y. Then there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon, d_Y) \subseteq N$. By Theorem 2.13.4, we have the equality: $S(x, \varepsilon, d_Y) = S(x, \varepsilon, d) \cap Y$. Take $N^* = N \cup S(x, \varepsilon, d)$. The converse part is obvious.

Corollary 2.13.8 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). Every neighbourhood in Y of a point x in Y is a neighbourhood of x in X if and only if Y is a neighbourhood of x in X.

Theorem 2.13.9 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). A set F is closed in Y if and only if there exists a closed set F^* in X such that $F = F^* \cap Y$.

Proof. Let F be a closed set in Y. Then $Y - F$ is an open set in Y. So there exists an open set G^* in X such that $Y - F = G^* \cap Y$ by Theorem 2.13.5. Then, $F = (X - G^*) \cap Y$, and the necessity part follows. The converse part is left to the reader.

Corollary 2.13.10 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). Every closed set in Y is a closed set in X if and only if Y is a closed set in X.

Theorem 2.13.11 Let (X, d) be a metric space, and (Y, d_Y) be a subspace of (X, d). Then the closure of a subset A of Y in Y is equal to $\overline{A} \cap Y$, where $\overline{A}$ is the closure of A in X.

Proof. Let x be a point in Y, and N be a neighbourhood of x in X. Then $N \cap A = (N \cap Y) \cap A$, and the result follows by Theorem 2.13.7.

Theorem 2.13.12 Let (X, d) be a metric space, and Y be a dense subset of X. Then for every point x in Y and every neighbourhood N of x in Y, the closure $\overline{N}$ of N in X is a neighbourhood of x in X.

Proof. By the definition of a neighbourhood, it follows that there is an open neighbourhood U in X of x such that $U \cap Y \subseteq N$. We show that $U \subseteq \overline{N}$. Let y be any point in U, and V be any neighbourhood of y in X. Then $U \cap V$ is a neighbourhood of y in X. Since Y is dense in X, $(U \cap V) \cap Y$ is not empty. This means that $(U \cap Y) \cap V$ is not empty. Therefore, $y \in \overline{(U \cap Y)} \subseteq \overline{N}$. Since U is an open neighbourhood of x, and $U \subseteq \overline{N}$, $\overline{N}$ is a neighbourhood of x in X.

Corollary 2.13.13 Every subspace of a separable metric space is separable.

Proof. Let (X, d) be a separable metric space, and Y be a subspace of X. Let $\mathbb{B}$ be a countable open base for X. Then the set $\mathbb{B}_Y = \{V \cap Y : V \in \mathbb{B}\}$ is a countable set. We show that $\mathbb{B}_Y$ is an open base for Y. Indeed, the sets in $\mathbb{B}_Y$ are open in Y by Theorem 2.13.5. Let B be an open set in Y. Then there exists an open set U in X such that $B = U \cap Y$. Since $\mathbb{B}$ is an open base for X, there exists a subcollection $\mathbb{B}_0$ of $\mathbb{B}$ such that $U = \bigcup\{V : V \in \mathbb{B}_0\}$. Now we have

$$B = U \cap Y = \bigcup\{V \cap Y : V \in \mathbb{B}_0\}.$$

Thus, $\mathbb{B}$ is the union of a subcollection of $\mathbb{B}_Y$. Therefore, Y is separable.

Theorem 2.13.14 Let (X, d) be a metric space, and Y be a separable subspace of X. If $\overline{Y} = X$, then (X, d) is separable.

Proof. Let A be a countable and dense subset of Y. Let $\varepsilon > 0$ be a real number, and x be a point in X. Since $\overline{Y} = X$, there exists a point y in Y such that $d(x, y) < \varepsilon/2$. Since A is a dense in Y, there exists a point z in A such that $d_Y(y, z) < \varepsilon/2$. Since $d_Y(y, z) = d(y, z)$, the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$ shows that $d(x, z) < \varepsilon$. Therefore, A is dense in X.

Definition 2.13.15 A property of a metric space (X, d) is said to be hereditary if every subspace X has that property.

The property of separability is hereditary. The Hausdorff property of Theorem 2.5.10 is also hereditary.

Theorem 2.13.16 Every uncountable subset of a second countable metric space contains uncountably many limit points of itself.

Proof. Let Y be a uncountable subset of a second countable metric space. Y is separable by Corollary 2.13.13. Therefore, there exists a countable and dense subset A of Y. Each point of Y is either a point of A or a limit point of A. But $Y - A$ is uncountable. Since each point of $Y - A$ is a limit point of A, Y has uncountably many limit point of itself.

The converse of the above theorem is also true.

Theorem 2.13.17 If (X, d) is a metric space in which every uncountable subset of X has a limit point, then X is second countable.

Proof. In view of Theorem 2.12.4, it is enough to show that X is separable. To show that X is separable, the scheme of the proof is to choose, for each positive n greater than or equal to a positive integer n_0 a maximal set A_n with the property that $d(x, y) \geq 1/n$ for all

$x, y \in A_n$, and for every point x in $X - A_n$, there is a point y in A_n such that $d(x, y) < 1/n$. For such a set A_n, obviously, no point of A_n is a limit point of A_n, and hence A_n can not be uncountable by our assumption. Then the union $\bigcup_{n=n_0}^{\infty} A_n$ is a countable and dense set in X.

The sets A_n are chosen as follows: Let n be a positive integer. Consider all sets A_α having the property that for any distinct points x, y in A_∞, the distance between x and y is at least $1/n$. Such sets obviously exist for sufficiently large values of n. Partially order this collection $\mathcal{A}$ of A_α's by set inclusion. Then starting with any fixed A_α, there is a maximal totally ordered subcollection $\{A_\beta\}$ such that each A_β contains A_α, by the maximal principle. Let $A_n = \bigcup_\beta A_\beta$. Then A_n is also some A_β. If x and y are any distinct points in A_n, then x and y are the elements of some A_β, and so $d(x, y) \geq 1/n$. Also A_n includes each A_β. Now if there is a point x of $X - A_n$ such that $d(x, y) \geq 1/n$ for all y in A_n, then $A_n \cup \{x\}$ is some A_β containing A_n and this contradicts the maximality of A_n. The proof is now complete.

Example 2.13.18 The metric space (X, d) of all bounded real (or complex) sequences in Example 2.1.23 is not separable. Let

$$S = \{\langle x_n \rangle : x_n = 0 \text{ or } 1 \text{ for all } n \in \mathbb{Z}_+\}.$$

Then S is uncountable. Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be any two distinct sequences in S. Then there exists a $k \in \mathbb{Z}_+$ such that $x_k \neq y_k$. Then $d(\langle x_n \rangle, \langle y_n \rangle) = 1$. So no point of S can be a limit point of S. By Theorem 2.13.16, X is not separable.

EXERCISE

1. Prove that every infinite metric space contains a denumerable discrete subspace.
2. Let CS be the set $\{\langle x_n \rangle : \langle x_n \rangle$is a real sequence such that the infinite series $\sum_n x_n$ is convergent$\}$ and $d : CS \times CS \to \mathbb{R}$ be defined by

$$d(x, y) = \sup_n |\sum_{i=1}^{n} (x_i - y_i)|$$

for all $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in CS. Show that d is a metric on CS. Further show that (CS, d) is a subspace of the sequence space

$$BS = \{\langle x_n \rangle : \langle x_n \rangle \text{ is a real sequence such that } \sup_n |\sum_{i=1}^{n} x_i| < \infty\}$$

with the metric ρ $\qquad \rho(x, y) = \sup_n |\sum_{i=1}^{n} (x_i - y_i)|$

for all $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in BS.

2.14 Pseudometric Spaces

Definition 2.14.1 Let X be a non-empty set, $d : X \times X \to \mathbb{R}$ be a mapping which satisfies the following properties:

(i) $d(x, y) \geq 0$ for all x, y in X,

(ii) $x = y$ implies that $d(x, y) = 0$ for all x, y in X,

(iii) $d(x, y) = d(y, x)$ for all x, y in X,

(iv) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x,\ y$ and z in X.

Then d is called a pseudometric on X, and (X, d) is called a pseudometric space.

If d is a metric on a set X, then d is a pseudometric on X, but not conversely.

Examples 2.14.2

1. Let X be a non-empty set. Then the mapping $d : X \times X \to \mathbb{R}$ defined by $d(x, y) = 0$ for all x, y in X is a pseudometric on X. Thus, (X, d) is a pseudometric space. Note that (X, d) is not a metric space, if X contains more than one points.

2. Let (X, d) be a metric space, and $\mathbb{B}_0$ be the collection of all bounded subsets of X. Let $\tau : \mathbb{B}_0 \times \mathbb{B}_0 \to \mathbb{R}$ be a mapping defined by

   $\tau(A, B) = \sup\{d(x, B) : x \in A\}$ for all A, B in $\mathbb{B}_0$.

   (See Example 2.4.7). Let us define a mapping $\sigma : \mathbb{B}_0 \times \mathbb{B}_0 \to \mathbb{R}$ by

   $$\sigma(A, B) = \max\{\tau(A, B), \tau(B, A)\}.$$

   Then σ is non-negative mapping. Also $\sigma(A, A) = 0$. σ is symmetric, that is, $\sigma(A, B) = \sigma(B, A)$. In order to prove triangle inequality, let A, B and C be sets in $\mathbb{B}_0$. Then, by Example 2.4.7

   $$\tau(A, B) \leq \tau(A, C) + \tau(C, B) \leq \sigma(A, C) + \sigma(C, B).$$

   Interchanging A and B, we have

   $$\tau(B, A) \leq \tau(B, C) + \tau(C, A) \leq \sigma(B, C) + \sigma(C, A).$$

   Using symmetry, the triangle inequality : $\sigma(A, B) \leq \sigma(A, C) + \sigma(C, B)$ follows. Thus, σ is a pseudometric on $\mathbb{B}_0$. This example represents a process of obtaining a pseudometric from a given metric space.

Corresponding to a pseudometric space (X, σ), there is a natural way to generate a metric space. The following theorem is a step in that direction.

Theorem 2.14.3 Let (X, d) be a pseudometric space. Then the relation $\sim$ on X defined by $x \sim y$ if $d(x, y) = 0$ is an equivalence relation.

Proof. Since $d(x, x) = 0,\ x \sim x$. Thus, $\sim$ is reflexive. The fact that $d(x, y) = d(y, x)$ together with $x \sim y$ implies that $y \sim x$, so the relation $\sim$ is symmetric. If $x \sim y$ and $y \sim z$, then the fact $x \sim z$ follows using the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$. Therefore, the relation $\sim$ is an equivalence relation on X.

The equivalence relation $\sim$ in the above theorem thus partitions the set X into equivalence classes. Denoting the equivalence class of an element x by $[x]$, and by $X/\sim$ the set of all equivalence classes with respect to the relation $\sim$, the following theorem makes sense.

Theorem 2.14.4 Let (X, d) be a pseudometric space. Then $(X/\sim, \rho)$, where $\rho([x], [y]) = d(x, y)$ for all x, y in X, is a metric space.

Proof. Since $x_1 \in [x]$ and $y_1 \in [y]$ implies that $d(x_1, y_1) = d(x, y)$, ρ is well defined. Obviously ρ is a pseudometric since d is a pseudometric. So it remains to show that $\rho([x], [y]) = 0$ implies that $[x] = [y]$. But then $\rho([x], [y]) = 0$ implies that $d(x, y) = 0$. So $x \sim y$, and $[x] = [y]$.

In Example 2.14.2(2), if $\sigma(A, B) = 0$, then it follows that $\overline{A} = \overline{B}$. On the other hand, if $\overline{A} = \overline{B}$, then $\sigma(A, B) = 0$. Hence, if $\mathcal{H}$ is the set of all closed and bounded sets in X, then $\sigma_0 = \sigma|_{(\mathcal{H}\times\mathcal{H})}$ is a metric on $\mathcal{H}$ called the Hausdorff metric on $\mathcal{H}$. The hausdorff metric σ_0 on the collection $\mathcal{H}$ of bounded and closed sets has been derived from the metric d on X, and it is thus indicative of a process by which the metrics on collections of subsets are obtained from the metric of the parent space.

EXERCISE

1. Let X be the set of all Riemann integrable real valued functions defined on $[a, b]$, and $d(f, g) = \int_a^b |f(t) - g(t)|dt$ for all f, g in X. Then (X, d) is a pseudometric space.
2. Let X be the set of all real valued differentiable functions defined on $[a, b]$, and $d(f, g) = \max\{|f'(t) - g'(t)| : t \in [a, b]\}$, where f' and g' denote the derivative of f and g, respectively. Show that (X, d) is pseudometric space.
3. Let (X, d) be a pseudometric space, and let $\rho(x, y) = \min\{1, d(x, y)\}$. Show that (X, ρ) is a pseudometric space.
4. Let X be a non-empty set, and $d : X \times X \to \mathbb{R}$ be a mapping such that
   (a) $d(x, y) \leq d(x, z) + d(y, z)$, and
   (b) $d(x, x) = 0$ for all x, y and z in X. Show that $d(x, y) \geq 0$, and $d(x, y) = d(y, x)$ for all x, y in X, that is, any d satisfying (a) and (b) is a pseudometric.
5. Let X be any non-empty set. Let $d : X \times X \to \mathbb{R}$ be a mapping such that $d(x, y) = 0$ for all x, y in X.

   Show that d is a pseudometric on X. Further show that unless X is a singleton, d is not a metric on X. Such a d is called trivial pseudometric on X.
6. Let (X, d) be a discrete metric space, and σ be the Hausdorff metric on the set $\mathcal{H}$ of closed sets in X. Prove that $\sigma(A, B) = 1$ for all $A \neq B$.
7. Let $A = \{(0, x) : -1 \leq x \leq 1\}$, and $B = \{(x, 0) : -1 \leq x \leq 1\}$ be sets in $\mathbb{R}^2$, and σ be the Hausdorff metric on the set $\mathcal{H}$ of closed and bounded sets. Find $\sigma(A, B)$.
8. (i) Let (X, d) be a pseudometric spae, and $r > 0$ be a real number. Let A be a non-empty subset of X. Define $D_r(A) = \cup\{S(x, r) : x \in A\}$ (See Example 2.4.2(8).) Let A, B and C be non-empty subsets. If $A \subseteq D_s(B)$, and $B \subseteq D_t(C)$, then prove that $A \subseteq D_{s+t}(C)$.
   (ii) Let (X, d) be a metric space, and $\mathcal{H}$ be a set of closed and bounded sets in X. Then $\sigma : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ defined by
   $$\sigma(A, B) = \inf\{r : A \subseteq D_r(B), \text{and } B \subseteq D_r(A)\}$$
   is a metric on $\mathcal{H}$ which is same as σ_0 defined in the last paragraph of this section.
   (iii) For any A and B in $\mathcal{H}$, and $r > 0$, prove that $\sigma(A, B) < r$ if and only if $A \subseteq D_r(B)$, and $B \subseteq D_r(A)$.
9. Let Y be a subspace of a metric space (X, d). Let $A \subseteq Y, A_Y^0$, be the interior of A in Y, and A_X^0 be the interior of A in X. Show that $A_X^0 \subseteq A_Y^0$, and give an example to show that the two may not be equal.

2.15 Product of Metric Spaces

From the several given metric spaces, we can form a new metric space known as the product metric space. The following theorem is the basis for defining such a metric space.

Theorem 2.15.1 Let $(X_i, d_i), i = 1, 2, \ldots, n$ be metric spaces, and $X = \prod_{i=1}^{n} X_i$. Then the mapping $d : X \times X \to \mathbb{R}$ defined by

$$d(x, y) = \max\{d_i(x_i, y_i) : i = 1, 2, \ldots, n\}$$

for all $x = (x_1, x_2, \ldots, x_n)$, and $y = (y_1, y_2, \ldots, y_n)$ in X is a metric on X.

Proof. As usual we verify the triangle inequality leaving the other properties of a metric to the reader. Let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, and $z = (z_1, z_2, \ldots, z_n)$ be any elements in X. Then for each $i = 1, 2, \ldots, n$,

$$\begin{aligned} d_i(x_i, z_i) &\leq d_i(x_i, y_i) + d_i(y_i, z_i) \\ &\leq \max_i\{d_i(x_i, y_i)\} + \max_i\{d_i(y_i, z_i)\} \\ &= d(x, y) + d(y, z). \end{aligned}$$

Therefore, $\max_i\{d_i(x_i, z_i)\} \leq d(x, y) + d(y, z)$, that is, $d(x, z) \leq d(x, y) + d(y, z)$. Thus, d is a metric on X, and (X, d) is a metric space.

The metric space (X, d) in the above theorem is known as product metric space, and we write it explicitly in the form of a definition.

Definition 2.15.2 Let (X_i, d_i), $i = 1, 2, \ldots, n$ be metric spaces, and $X = \prod_{i=1}^{n} X_i$. Then the mapping defined by

$$d(x, y) = \max_i\{d_i(x_i, y_i)\}$$

for all $x = (x_1, x_2, \ldots, x_n)$, and $y = (y_1, y_2, \ldots y_n)$ in X is a metric on X. The metric space (X, d) is known as the product metric space of the metric spaces (X_i, d_i), $i = 1, 2, \ldots, n$ which are often called the factors of the product.

Once the product metric space is defined, it is relevant to know the structure of open spheres, open sets, closed sets, limit points, closures etc in this metric space.

Theorem 2.15.3 Let (X_i, d_i), $i = 1, 2, \ldots, n$ be metric spaces, and (X, d) be their product metric space. Then for each $x = (x_1, x_2, \ldots, x_n)$ in X and a real number $\varepsilon > 0$,

$$S(x, \varepsilon, d) = S(x_1, \varepsilon, d_1) \times S(x_2, \varepsilon, d_2) \times \ldots \times S(x_n, \varepsilon, d_n).$$

Proof. The proof is left to the reader.

Theorem 2.15.4 Let (X_i, d_i), $i = 1, 2, \ldots, n$ be metric spaces, and (X, d) be their product metric space. If for each $i = 1, 2, \ldots, n$, A_i is an open set in X_i, then $A_1 \times A_2 \times \ldots \times A_n$ is an open set in X.

Proof. For any point $x = (x_1, x_2, \ldots, x_n)$ in $A_1 \times A_2 \times \ldots \times A_n$, $x_i \in A_i$ for all $i = 1, 2, \ldots, n$. Since A_i's are open sets in X_i, there exists a real number ε_i such that $S(x_i, \varepsilon_i, d_i) \subseteq A_i$ If $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$, then, by Theorem 2.15.3, it follows that $S(x, \varepsilon, d) \subseteq A_1 \times A_2 \times \ldots \times A_n$, and the proof is complete.

If (X, d) is the product metric space of the metric spaces (X_i, d_i), $i = 1, 2, \ldots, n$, and G is an open set in X, then the projection of G on the i^{th} coordinate $\pi_i(G) = \{x_i :$

$x = (x_1, x_2, \ldots, x_i, \ldots, x_n) \in G\}$ is an open set in X_i for all $i = 1, 2, \ldots, n$. To see this, let $x_i \in \pi_i(G)$. Then, there exists a point $x = (x_1, x_2, \ldots, x_i, \ldots, x_n)$ in G such that its projection on the i^{th} co-ordinate $\pi_i(x) = x_i$, and there is a real number $\varepsilon > 0$ such that $S(x, \varepsilon, d) \subseteq G$. Then, by Theorem 2.15.3, $\pi_i(S(x, \varepsilon, d)) = S(x_i, \varepsilon, d_i)$ which is contained in $\pi_i(G)$. Therefore $\pi_i(G)$ is an open set in X_i for all $i = 1, 2, \ldots, n$.

The following theorem expresses the interior (closure) of a set in a product metric space as the product of the interiors (closures) of the factor sets in the factor spaces.

Theorem 2.15.5 Let $(X_i, d_i),\ i = 1, 2, \ldots, n$ be metric spaces, and (X, d) be their product metric space. For each $i = 1, 2, \ldots, n$, let A_i be a subset in X_i. Then

$$(1)\ \left(\prod_{i=1}^{n} A_i\right)^0 = \prod_{i=1}^{n} A_i^0, \text{ and } (2)\ \overline{\left(\prod_{i=1}^{n} A_i\right)} = \prod_{i=1}^{n} \overline{A}_i,$$

where $\left(\prod_{i=1}^{n} A_i\right)^0$ denotes the interior of $\prod_{i=1}^{n} A_i$ in X and A_i^0 is the interior of A_i in X_i; similarly, in (2).

Proof. (1) A point $x = (x_1, x_2, \ldots, x_n) \in \left(\prod_{i=1}^{n} A_i\right)^0$ if and only if there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon, d) \subseteq \left(\prod_{i=1}^{n} A_i\right)$ if and only if $S(x, \varepsilon, d_i) \subseteq A_i$ for all $i = 1, 2, \ldots, n$ by Theorem 2.15.3 if and only if $x_i \in A_i^0$ for all $i = 1, 2, \ldots, n$ if and only if $x \in \prod_{i=1}^{n} A_i^0$.

(2) A point $x = (x_1, x_2, \ldots, x_n) \in \overline{\left(\prod_{i=1}^{n} A_i\right)}$ if and only if $S(x, \varepsilon, d) \cap \left(\prod_{i=1}^{n} A_i\right) \neq \emptyset$ for each real number $\varepsilon > 0$ if and only if $\prod_{i=1}^{n} S(x_i, \varepsilon, d_i) \cap \prod_{i=1}^{n} A_i \neq \emptyset$ for each real number $\varepsilon > 0$ if and only if $S(x_i, \varepsilon, d_i) \cap A_i \neq \emptyset$ for each real number $\varepsilon > 0$, and $i = 1, 2, \ldots, n$ if and only $x \in \prod_{i=1}^{n} \overline{A}_i$.

Theorem 2.15.6 Let $(X_i, d_i),\ i = 1, 2, \ldots, n$ be metric spaces, and (X, d) be their product metric space. Let $A_i \subseteq X_i$, for all $i = 1, 2, \ldots, n$ Then $A_i \times A_2 \times \ldots \times A_n$ is closed in X if and only if A_i is closed in X_i for all $i = 1, 2, \ldots, n$.

Proof. Let $A = A_1 \times A_2 \times \ldots \times A_n$. Then A is closed in X if and only if $\overline{A} = A$ if and only if $\prod_{i=1}^{n} \overline{A}_i = \prod_{i=1}^{n} A_i$ if and only if $\overline{A}_i = A$ for all $i = 1, 2, \ldots, n$ if and only if A_i is a closed set for all $i = 1, 2, \ldots, n$.

For the following theorem, we refer to Example 2.1.14.

Theorem 2.15.7 Let $(X_i, d_i),\ i = 1, 2, \ldots, n$ be metric spaces, and (X, d) be their product metric space. Then the metrics

$$d^*(x, y) = \left[\sum_{i=1}^{n} \{d_i(x_i, y_i)\}^2\right]^{1/2},$$

and $d^{**}(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i)$

for all $x = (x_1, x_2, \ldots, x_n)$, and $y = (y_1, y_2, \ldots, y_n)$ in X are both equivalent to d.

Proof. Observe that

$$d(x, y) \leq d^*(x, y) \leq \sqrt{n} d(x, y),$$
$$\text{and} \qquad d(x, y) \leq d^{**}(x, y) \leq n d(x, y)$$

for all x, y in X, and apply Theorem 2.6.6.

So far in this section, we have considered only the product of finite number of metric spaces. Now consider the product of infinite number of metric spaces. In Examples 2.1.8–2.1.9 we defined the metrics d_p and d_∞ on $\mathbb{R}^n$. As a first attempt, these metrics are extended to $\mathbb{R}^\omega$ by defining

$$d_p(x,y) = \left[\sum_{i=1}^{\infty}(x_i - y_i)^p\right]^{1/p}, \ 1 \le p < \infty$$

and $d_\infty(x,y) = \sup\{|x_i - y_i| : i \in \omega\}$.

But these definitions simply do not make sense; In the first case, the series may not be convergent, and in the second case the sup may not exist. However, if we take the standard bounded metric on $\mathbb{R}$ (see Example 2.1.16) given by $\overline{d}(x,y) = \min\{1, |x-y|\}$, and define $\rho : \mathbb{R}^\omega \times \mathbb{R}^\omega \to \mathbb{R}$ by

$$\rho(x,y) = \sup\{\overline{d}(x_i, y_i)\}$$

for all $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2 \ldots)$ in $\mathbb{R}^\omega$, then it is easy to see that ρ is a metric on $\mathbb{R}^\omega$. This illustration is a particular case of the more general result in the following theorem.

Theorem 2.15.8 Let $\{(X_n, d_n) : n \in \omega\}$ be a countably infinite family of metric spaces, and $X = \prod_{i\in\omega} X_i$. Let $d = X \times X \to \mathbb{R}$ be the mapping defined by

$$d(x,y) = \sup\{d_n(x_n, y_n) : n \in \omega\}$$

for all $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ in X. Then d is a metric on X if diam $(X_n) \le M$, where M is a real number, for all large n.

Proof. Since $\text{diam}(X_n) \le M < \infty$ for some fixed constant real number M for all $n \ge n_0$, where n_0 is some positive integer, it follows that $d(x,y)$ is well-defined. d clearly satisfies the first three properties of a metric, and the triangle inequality is proved as follows :

$$\begin{aligned} d(x,z) &= \sup_n\{d_n(x_n, z_n)\} \le \sup_n\{d_n(x_n, y_n)\} + \sup_n\{d(y_n, z_n)\} \\ &= d(x,y) + d(y,z). \end{aligned}$$

It may be noted that there is nothing special about the set ω in $\mathbb{R}^\omega$, as far as the metric ρ in the illustration preceding to the above theorem is concerned. We can define it for $\mathbb{R}^J$ for arbitrary J as follows.

Definition 2.15.9 Let J be any arbitrary index set, and $\{(X_\alpha, d_\alpha)\}_{\alpha\in J}$ be a family of metric spaces. Let $X = \prod_{\alpha\in J} X_\alpha$ and $\overline{d}_\alpha(x_\alpha, y_\alpha) = \min\{1, d_\alpha(x_\alpha, y_\alpha)\}$. Let the mapping $\overline{\rho} : X \times X \to \mathbb{R}$ be defined by

$$\overline{\rho}(x,y) = \sup_\alpha\{\overline{d}_\alpha(x_\alpha, y_\alpha)\}$$

for all $x, y \in X$ which are defined by $x(\alpha) = x_\alpha \in X_\alpha$ and $y(\alpha) = y_\alpha \in X_\alpha$ for all $\alpha \in J$. Then $\overline{\rho}$ is a metric on X called the uniform metric.

There is another metric on a countable product, which is in fact known as product metric.

Theorem 2.15.10 Let $\{(X_i, d_i) : i \in \omega' \subseteq \omega\}$ be a countable family of metric spaces. Let $X = \prod_{i\in\omega'} X_i$. Let $D : X \times X \to \mathbb{R}$ be the mapping defined for $x = (x_i)$ and $y = (y_i)$ in X by

$$D(x,y) = \text{lub}_{i\in\omega'}\{\rho_i(x_i, y_i)/i\}$$

where $\rho_i(x_i, y_i) = \min\{d_i(x_i, y_i), 1\}$ is the standard bounded metric on X_i, and $i \in \omega$. Then D is a metric on X known as a product metric on X.

Proof. The triangle inequality for D follows from the following observation :

$$\rho_i(x_i, y_i)/i \leq \rho_i(x_i, z_i)/i + \rho_i(z_i, y_i)/i$$
$$\leq D(x, z) + D(z, y),$$

and so $$\text{lub}_{i\in\omega'}\{\rho_i(x_i, y_i)/i\} \leq D(x, z) + D(y, z).$$

The following theorem refers to the product metric D of Theorem 2.15.10, the uniform metric $\overline{\rho}$ in Definition 2.15.9 and the product metric d in Definition 2.15.2.

Theorem 2.15.11 On finite products, the product metric D, the uniform metric $\overline{\rho}$ and the product metric d are equivalent.

Proof. Let $(X_i, d_i),\ i = 1, 2, \ldots, n$ be metric spaces, and $X = \prod_{i=1}^{n} X_i$. Observe that

$$S(x, \varepsilon/n, D) \subseteq S(x, \varepsilon, \overline{\rho}) \subseteq S(x, \varepsilon, D).$$

Also for any real number $\varepsilon > 0,\ S(x, \varepsilon, d) \subseteq S(x, \varepsilon, \overline{\rho})$. Further, if $\varepsilon^* > 0$ be any real number, then there exists a real number ε' such that $0 < \varepsilon' < 1$, and $S(x, \varepsilon', \overline{\rho}) \subseteq S(x, \varepsilon^*, d)$. Now, by Theorem 2.6.4, it follows that $D, \overline{\rho}$ and d are equivalent.

EXERCISE

1. For each $i \in \mathbb{Z}_+$, let $X_i = \{0, 2\}$, and X_i be equipped with usual metric. Let $X = \prod_{i=1}^{\infty} X_i$, and d be the metric on X, defined by

$$d(x, y) = \sum_{i-1}^{\infty} |x_i - y_i|/2^i[1 + |x_i - y_i|]$$

where $x = \langle x_i \rangle$ and $y = \langle y_i \rangle$ are any points in X. Show that the metric space (X, d) is a subspace of the real Fréchet metric space in Example 2.1.22.

2.16 The Cantor Set

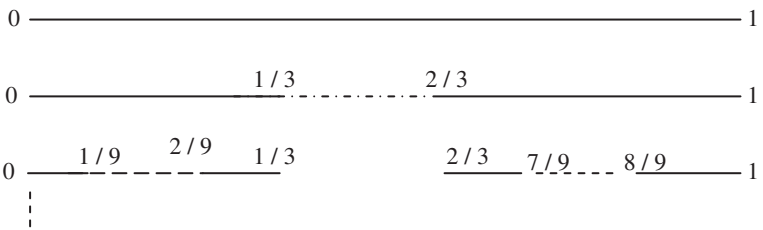
We have already seen that a closed interval in the real line $\mathbb{R}$ is a closed set in $\mathbb{R}$. The union of finite number of such closed intervals is a closed set by Theorem 2.7.3. These and other closed sets in $\mathbb{R}$ are characterised by Theorem 2.5.9: the closed sets in $\mathbb{R}$ are the complements of unions of countable disjoint classes of open intervals. The cantor set is also defined similarly as the complement of the union of a countable class of disjoint open intervals.

The intuitive idea of the cantor set is rather simple. Remove from $\mathbb{R}$ the open intervals $]-\infty, 0[$ and $]1, \infty[$. Then we are left with the closed interval $[0, 1]$, denoted by F_0. In the next step, delete from F_0 the middle third open interval $]1/3, 2/3[$. Then we are left with the closed set $F_1 = [0, 1/3] \cup [2/3, 1]$. Again similarly from the two intervals in F_1, delete

their middle third open intervals, that is, delete $]1/9, 2/9[$ from $[0, 1/3]$ and $]7/9, 8/9[$ from $[2/3, 1]$. The remaining closed set $F_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$. Continuing this process of deleting the middle third open interval from each of the closed intervals left from the previous step, we obtain a sequence $\langle F_n \rangle$ of closed sets in $\mathbb{R}$ such that F_{n+1} is properly contained in F_n. The cantor set C is defined by

$$C = \bigcap_{n=1}^{\infty} F_n$$

The Cantor set C is a closed set by Theorem 2.7.3. The geometrical illustration of C is given in Fig. 2.8



The Cantor set C

Fig. 2.8

The cantor set in a closed interval $[a, b]$, $a < b$, in $\mathbb{R}$ can be defined similarly as above which in fact is an inductive definition:

Set $F_0 = [a, b]$. Suppose F_n is defined as the union of finite number of disjoint closed intervals, that is, suppose $F_n = I_1 \cup I_2 \cup \ldots \cup I_m$ in which each I_j is a closed interval and $I_j \cap I_k$ is empty whenever $j \neq k$. Then F_{n+1} is defined as follows: first delete from each $I_j, 1 \leq j \leq m$, its middle third open interval; the two closed subintervals left after deleting the open interval are denoted by I_{j_1} and I_{J_2}; Now $F_{n+1} = I_{1_1} \cup I_{1_2} \cup \ldots \cup I_{m_1} \cup I_{m_2}$. Now, by induction, F_n is defined for all n. The Cantor set C is now defined by $C = \bigcap_{n=1}^{\infty} F_n$. Since each F_n is a closed set, C is a closed set.

Note that F_n contains 2^n disjoint closed subintervals of $[0, 1]$, and there are 2^{n+1} end points of these subintervals. All these points are the elements of C for each n. Are there other points in C? The set C is not a finite set. For any $k \in \mathbb{Z}_+$, F_k contains 2^k closed subintervals, and so 2^{k+1} end points of these subintervals are in C, and $k < 2^{k+1}$ for all k. Thus, C is an infinite set. There are multitudes of other points in C which are not the end points of the intervals in F_k for any k. For instance, $1/4$ is a point in C, and it is not an end point of any subinterval in F_k for any k.

Let us recall that the ternary expansion of a decimal number a in $[0, 1]$ is of the form $a = \cdot a_1 a_2 \ldots$, where each $a_i = 0, 1$ or 2. The decimal equivalent of this number is given by

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$$

It is easy to see that the ternary expansion of a number is not unique. For example, the ternary expansions

$$x = .00 \ldots 0100 \ldots$$

$$y = .00 \ldots 0022 \ldots$$

where x has 1 at the q^{th} place and 0 elsewhere, and y has zeros upto q^{th} place and two's after that are both equivalent to the decimal number $1/3^q$. However, if we consider the ternary number of the form $.a_1a_2\ldots$ in which each a_i is either 0 or 2, then any two distinct such ternary numbers represent different decimal numbers. The cantor set C consists of exactly these later numbers, that is, numbers in $[0,1]$ whose ternary expansion contains only zero's and two's. Consider a ternary number $x = .a_1a_2\ldots$ in $[0,1]$ consisting of digits zero's and two's. Denote the subinterval $[0,1/3]$ by I_{11}, and the interval $[2/3,1]$ by I_{12} so that $F_1 = I_{11} \cup I_{12}$. Now if the digit a_1 in x is $0, x$ is in I_{11} otherwise x is in I_{12}. Rename the constituent interval of F_1 containing x as I_1. Then F_2 contains the two closed subintervals of I_1 that are left after deleting the middle third open interval of I_1; denote these subintervals by I_{21} and I_{22} (Fig. 2.9).

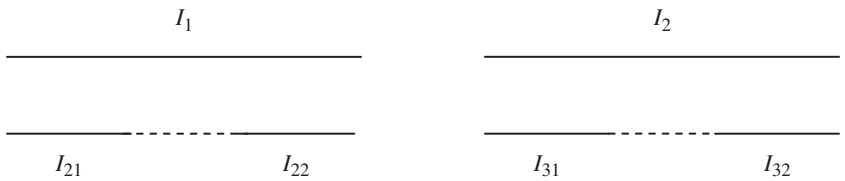


Fig. 2.9

Now if the digit a_2 in x is 0, then x is in I_{21} otherwise x is in I_{22}. Rename the constituent interval of I_1 containing x as I_2, and by continuing the above process, it is now immediate that x is in C. Since $1/4 = \sum_{n=1}^{\infty} 2/3^{2n}$, it has the ternary expansion $.020202\ldots$. Thus, $1/4$ is in C. Conversely, if x is a point in C, then it has a ternary expansion consisting of the zero's and two's. Since x is in C, it is in F_n for all n. Since x is in $F_1 = I_{11} \cup I_{12}$, then the first ternary digit is 0 if x is in I_{11} and 2 if x is in I_{12}. Rename the subinterval of F_1 containing x as I_1 as above. Then, the second digit of x is 0 if x is in I_{21}, and 2 if x is in I_{22}, and so on. It now follows that the ternary expansion of x consists of zeros and twos. For this reason the set C is sometimes called the Cantor ternary set.

C is, in fact, an uncountable set. We define a mapping $f : [0,1[\to C$ as follows. Assume that $.a_1a_2\ldots$ is the binary expansion of a point x in $[0,1[$. Define $b_n = 2a_n$ for all n, and denote the ternary expansion, $.b_1, b_2, \ldots$ by y. Then y is in C. We define $f(x) = y$. This mapping f is the injective. Since $[0,1[$ is uncountable, it follows that C is uncountable, and has the same cardinality as that of continuum.

It is interesting to see that the sum of the lengths of the open intervals deleted to obtain the Cantor set is precisely $1 : 1/3 + 2/9 + 4/27 + \ldots = 1$. The Cantor set is the source of many examples in the theory of metric spaces. Beginners and mathematicians alike look at this mathematical object with awe and curiosity. We next give some additional facts about this set C, where it is assumed that C inherits the usual metric of $\mathbb{R}$.

1. The interior of the cantor set $C \subseteq \mathbb{R}$ is empty. Suppose on the contrary that x is an interior point of C. Then there exists a real number $\varepsilon > 0$ such that the open interval $]x-\varepsilon, x+\varepsilon[$ is contained in C. The diameter of the interval $]x-\varepsilon, x+\varepsilon[$ is 2ε. Also the diameter of each of the closed subintervals that constitute F_n is $1/3^n$ choose n large enough so that $1/3^n < \varepsilon$. Then the interval $]x-\varepsilon, x+\varepsilon[$ is not contained in any constituent subintervals of F_n. Hence, the interval $]x-\varepsilon, x+\varepsilon[$ is not contained in F_n. Thus, $]x-\varepsilon; x+\varepsilon[$ is not contained in C, a contradiction.

2. The Cantor set C is a nowhere dense set in $\mathbb{R}$. Since C is a closed set, $\overline{C} = C$. Therefore, $(\overline{C})^0 = C^0 = \emptyset$.

3. The Cantor set C is perfect. Since C is a closed set, it remains to show that C is dense-in-itself. Let $.a_1a_2\ldots$ be the ternary expansion of a point in C. Then the sequence $\langle x_n\rangle$ defined by
$$x_n = .a_1a_2\ldots a_n'a_{n+1}\ldots,$$
where $a_n' = 0$ if $a_n = 2$ and $a_n' = 2$ if $a_n = 0$, is in C, and it converges to $.a_1a_2\ldots$. Thus, C is dense in itself.

The Cantor sets have been defined in $\mathbb{R}^n$ too. We give here a brief sketch of their construction in $\mathbb{R}^2$. Let $F_0 = [0,1]$, and $J_0 = F_0 \times F_0$. Delete from J_0 the middle third star shaped open region, that is, the set $A_0 =]1/3, 2/3[\times[0,1]) \cup ([0,1]\times]1/3, 2/3[)$.

Denote by J_1 the remaining region, that is, $J_1 = J_0 - A_0$. Then

$J_1 = ([0,1/3]\times[0,1/3]) \cup ([0,1/3]\times[2/3,1]) \cup ([2/3,1]\times[0,1/3]) \cup ([2/3,1]\times[2/3,1])$
$= J_{1,1} \cup J_{1,2} \cup J_{1,3} \cup J_{1,4}$ (Fig. 2.10).

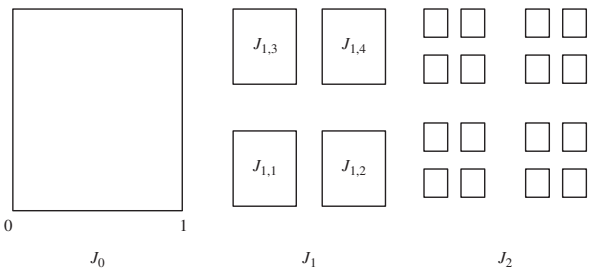


Fig. 2.10

Note that $J_{1,1}, J_{1,2}, J_{1,3}$ and $J_{1,4}$ are pairwise disjoint closed sets in $\mathbb{R}^2$. Therefore, J_1 is a closed set. Suppose that J_k is defined as a union of pairwise disjoint closed sets in $\mathbb{R}^2$ as $J_k = J_{k,1} \cup J_{k,2} \cup .. \cup J_{k,m}$. Then J_{k+1} is obtained from J_k is follows: delete from each $J_{k,i},\ i = 1, 2, \ldots, m$, its middle third star shaped open region; after deletion, the remaining part of $J_{k,i}$ consists of four disjoint rectangular closed regions in the set $\{J_{k,i_1}, J_{k,i_2}, J_{k,i_3}, J_{k,i_4}\}$. Now J_{k+1} is defined as the set

$J_{k+1} = J_{k,1_1} \cup J_{k,1_2} \cup J_{k,1_3} \cup J_{k,1_4} \cup J_{k,2_1} \cup J_{k,2_2} \cup J_{k,2_3} \cup J_{k,2_4} \cup \ldots.$
$\ldots \cup J_{k,m_1} \cup J_{k,m_2} \cup J_{k,m_3} \cup J_{k,m_4}.$

Redoing the suffixes,
$$J_{k+1} = J_{k+1,1} \cup J_{k+1,2} \cup \ldots \cup J_{k+1,4m}.$$

By induction, J_n is defined for $n \in \mathbb{Z}_+$. The planar Cantor set C is defined by
$$C = \bigcap_{n=1}^{\infty} J_n.$$

Since J_n is a closed set for each n, the planar Cantor set is a closed set.

EXERCISE

1. There is nothing important about deleting the middle third open intervals in the construction of the Cantor set C in $[0,1]$. We can delete the middle open interval of any diameter, less than the diameter of the closed interval under consideration during the process of obtaining F_n. Then the resulting set $S = \bigcap_{n=1}^{\infty} F_n$ will have similar properties as the Cantor set. As a special case, define

$$T_0 = [0,1]$$
$$T_1 = [0,1/4] \cup [3/4,1]$$
$$T_3 = [0,1/16] \cup [3/16,1/4] \cup [3/4,13/16] \cup [15/16,1]$$
$$\vdots$$

   Here T_{n+1} is obtained from T_n in the same manner as in the Cantor set: from each constituent closed subinterval I of T_n delete the middle open interval of diameter equal to half of the diameter of I. Set $S_{1/2} = \bigcap_{n=1}^{\infty} T_n$. Prove that each element x in $S_{1/2}$ has the quaternary form $x = 0.a_1a_2\ldots$, where each $a_i = 0$ or 3.

2. Let A be the set of all $x \in [0,1]$ whose decimal expansion contains only the digits 4 and 7. Is A countable? Is A dense in [0,1]? Is A closed? Is A perfect?

2.17 Totally Bounded Sets

The notion of a bounded set is a property of distance. There is another property called total boundedness which in addition to distance has something to do with the scatteredness of the elements in the metric space.

Definition 2.17.1 Let (X,d) be a metric space, and A be a subset of X. Let $\varepsilon > 0$ be a real number. Then a subset B of X is said to be an ε-net for A if for each point y in A, there exists a point x in B such that $d(x,y) < \varepsilon$.

Definition 2.17.2 A non-empty subset A of metric space is said to be totally bounded if for each real number $\varepsilon > 0$ there exists a finite ε-net for A.

The total boundedness is a stronger property than boundedness.

Theorem 2.17.3 Every totally bounded subset of a metric space is bounded.

Proof. Let (X,d) be a metric space, and A be a totally bounded subset of X. For $\varepsilon = 1$, let $B = \{x_1, x_2, \ldots, x_n\}$ be a finite ε-net for A. Now for each $x \in A$, there exists a point x_i in B such that $d(x,x_i) < \varepsilon$. Thus, $x \in S(x_i,\varepsilon)$, and so $A \subseteq \bigcup_{i=1}^{n} S(x_i,\varepsilon)$. Since an open sphere is a bounded set, and finite union of bounded sets is bounded, it follows that A is bounded.

The converse of the above theorem is not true. For example, any infinite subset A of an infinite discrete metric space is bounded, but there is no finite 1/2-net for A. Therefore, A is not totally bounded.

The following Lemma covers some equivalent conditions of total boundedness of sets.

Lemma 2.17.4 Let (X,d) be a metric space, and A be a non-empty subset of X. The following statements are equivalent.

1. A is totally bounded.
2. For each real number $\varepsilon > 0$, there exists a finite covering of A consisting of sets of diameter less than ε.
3. For each real number $\varepsilon > 0$, there exists a partition of A into sets of diameter less than ε.
4. For each real number $\varepsilon > 0$, there exists a finite ε-net for A consisting of points of A; In this case we say that finite ε-net is in A.

Proof. Let $\varepsilon > 0$ be a real number. Let B be a finite $\varepsilon/3$-net for A. Then the set $\{S(x, \varepsilon/3) : x \in B\}$ of open spheres is a finite cover of A, that is, $A \subseteq \bigcup_{x\in B} S(x, \varepsilon/3)$, consisting of sets of diameter less than ε. Therefore, (1) implies (2). Now let $\{A_1, A_2, \ldots A_n\}$ be a finite covering of A such that $\text{diam}(A_i) < \varepsilon$ for all $i = 1, 2, \ldots, n$. Then the collection $\mathbb{B}_1 = \{A_i \cap A_j : i \neq j\} \cup \{A_i - A_j : i \neq j\}$ is a covering of A. Delete from this the empty set, if any. Now repeat the process, and obtain the collection

$$\mathbb{B}_2 = \{B_i \cap B_j : B_i, B_j \in \mathbb{B}_1 \text{ and } i \neq j\} \cup \{B_i - B_j : B_i, B_j \in \mathbb{B}_1 \text{ and } i \neq j\}$$

Thus, $\mathbb{B}_2$ is also a covering of A. Since n is finite, in finite number of steps, we obtain a covering $\mathbb{B}_0$ of A consisting of disjoint sets. Now we take the intersection with A of each set in this covering $\mathbb{B}_0$, and deleting the empty set, if any, then the resulting set $\{B \cap A : B \in \mathbb{B}_0\} - \{\emptyset\}$ is a partition of A, and obviously each set in this partition is of diameter less than ε. The other implications are obvious.

Theorem 2.17.5 A bounded subset of $\mathbb{R}^n$ is totally bounded.

Proof. Let A be a bounded subset of $\mathbb{R}^n$, and $a = (a_1, a_2, \ldots, a_n)$ be a point in $\mathbb{R}^n$. Then there exists a real number $\varepsilon > 0$ such that A is contained in the open sphere $S(a, \varepsilon)$. Now $S(a, \varepsilon)$ is contained in the cube $]a_1 - \epsilon, a_1 + \epsilon[\times]a_2 - \varepsilon, a_2 + \varepsilon[\times \ldots \times]a_n - \varepsilon, a_n + \varepsilon[$ of edge 2ε. Each interval $[a_i - \varepsilon, a_i + \varepsilon]$, $i = 1, 2, \ldots, n$ is partitioned into m closed subintervals of equal length. We select a subinterval from the partition of each interval $[a_i - \varepsilon, a_i + \varepsilon]$, and form their product to get a cube of edge $2\epsilon/m$. There are m^n such sub-cubes each of which is of diameter $2\varepsilon\sqrt{n}/m$. (The diameter of a cube in $\mathbb{R}^n$ of edge l is $l\sqrt{n}$.) Thus, if $\varepsilon' > 0$ be a real number, then there exists an $m_0 \in \mathbb{Z}_+$ such that $2\varepsilon\sqrt{n}/m_0 < \varepsilon'$. Then the m_0^n sub-cubes of edge $2\varepsilon/m_0$ provide a finite covering of A, consisting of sets of diameter less that ε'. By Lemma 2.17.4, now A is totally bounded.

Theorem 2.17.6 A totally bounded metric space is separable, and therefore satisfies the second axiom of countability.

Proof. Let (X, d) be a totally bounded metric space. Then for each $n \in \mathbb{Z}_+$, there exists a finite $1/n$-net A_n for X. The set $M = \bigcup_{n=1}^{\infty} A_n$ is countable. We show that M is dense in X. Let x be any point in X. Let $\varepsilon > 0$ be a real number. Then there exists an $m \in \mathbb{Z}_+$ such that $1/m < \varepsilon$. Therefore, there exists a point y in the $1/m$-net A_m such that $d(x, y) < 1/m$, that is, $y \in S(x, \varepsilon)$. Thus, $S(x, \varepsilon) \cap M$ is non-empty, and $x \in \overline{M}$.

The converse of the above theorem does not hold. For instance, the real line $\mathbb{R}$ is separable, but not totally bounded.

Example 2.17.7 A closed unit sphere S in l_2-space is bounded but not totally bounded. Consider a sequence $\langle x_n \rangle$, where $x_n = \langle 0, 0, \ldots, 1, 0, \ldots \rangle$ consisting of 1 at the n^{th} place

and 0 at other places, in the unit sphere centered at the origin $\langle 0, 0, \ldots \rangle$. Then $d(x_n, x_m)$, where d is the metric of l_2-space, is equal to $\sqrt{2}$ for $n \neq m$. Now suppose that S is totally bounded, and $A = \{a_1, a_2, \ldots, a_k\}$ be a $1/\sqrt{2}$-net. Then one of the open spheres $S(a_i, 1/\sqrt{2})$, $i = 1, 2, \ldots, k$ must contain x_n for infinitely many values of n. Let x_n and x_m, $n \neq m$, be in $S(a_i, 1/\sqrt{2})$. Then

$$d(x_n, x_m) \leq d(x_n, a_i) + d(a_i, x_m) < 1/\sqrt{2} + 1/\sqrt{2} = \sqrt{2},$$

that is, $\sqrt{2} < \sqrt{2}$, a contradiction. Therefore, S is not totally bounded.

Example 2.17.8 The set $A = \{\langle p_i \rangle \in l_2\text{-space: } |p_i| \leq 1/2^i \text{ for all } i \in \mathbb{Z}_+\}$ is a totally bounded subset of l_2-space. To see this, let $\varepsilon > 0$ be a real number. We can choose $n \in \mathbb{Z}_+$ such that $1/2^n < \varepsilon/2$. We define another set

$$B_n = \{\langle q_i \rangle \in l_2\text{-space: there exists a } \langle p_i \rangle \text{ in } A \text{ such that}$$
$$q_i = p_i \text{ for all } i = 1, 2, \ldots, n, \text{ and } q_i = 0 \text{ for } i > n\}.$$

Then B_n is contained in A, and the distance between $\langle p_n \rangle$ in A and the corresponding $\langle q_n \rangle$ in B_n

$$d(\langle p_n \rangle, \langle q_n \rangle) = [\sum_{i=n+1}^{\infty} p_i^2]^{1/2} \leq [\sum_{i=n+1}^{\infty} (1/2^i)^2]^{1/2}$$
$$\leq [\sum_{i=n+1}^{\infty} 1/4^i]^{1/2} < 1/2^n < \varepsilon/2.$$

Now define the set $C_n = \{(q_1, q_2, \ldots, q_n) : \langle q_i \rangle \in B_n\}$. It is easy to see that the set C_n is a bounded subset $\mathbb{R}^n$, and therefore, C_n is totally bounded subset of $\mathbb{R}^n$. Therefore, there exists a finite $\varepsilon/2$-net A^* for C_n in C_n. Let B^* be the set of points $\langle x_1, x_2, \ldots, x_n, 0, 0, \ldots \rangle$ in l_2-space such that the point $(x_1, x_2, \ldots, x_n)$ is in A^*. Then $B^* \subseteq B_n$, and it follows that B^* is a finite $\varepsilon/2$-net for B_n, and a finite ε-net for A.

Theorem 2.17.9 The closure of a totally bounded subset of a metric space is totally bounded.

Proof. Let (X, d) be a metric space, and A be a totally bounded subset of X. Let $\varepsilon > 0$ be a real number. Then there exists a finite $\varepsilon/2$-net $\{x_1, x_2, \ldots x_n\}$ for A in A. Let x be a point in $\overline{A}$. Then the open sphere $S(x, \varepsilon/2)$ contains a point y in A. Since A is totally bounded, there exists a point x_i in the $\varepsilon/2$-net $\{x_1, x_2, \ldots, x_n\}$ such that y is in $S(x_i, \varepsilon/2)$. Therefore, $d(x, x_i) < \varepsilon$, and $\{x_1, x_2, \ldots, x_n\}$ is a finite ε-net for $\overline{A}$.

Theorem 2.17.10 A subspace of a totally bounded metric space is totally bounded.

Proof. Let (X, d) be a totally bounded metric space, and (Y, d) be a subspace of X. Since X is totally bounded, for positive real number ε, there exists a finite $\varepsilon/2$-net $A = \{x_1, \ldots, x_k\}$ in X. Let $\{x_{m_1}, \ldots, x_{m_l}\}$ be the subset of A consisting of all points whose distance from Y is less than $\varepsilon/2$, and let $x_1', x_2', \ldots, x_l'$ be arbitrary points of Y such that

$$d(x_{m_i}, x_i') < \varepsilon/2 \text{ for } i = 1, 2, \ldots, l.$$

We show that the set $A' = \{x_1', \ldots, x_l'\}$ is a finite ε-net in Y. Let x be any point in Y. Since A is a finite $\varepsilon/2$-net, there exists an $i \leq k$ such that $d(x, x_i) < \varepsilon/2$. Therefore, $x_i = x_{m_j}$ for some $j \leq l$, and so $d(x, x_j') < \varepsilon$.

The reader should observe that a metric space (X, d) is totally bounded if and only if $(X, \overline{d})$, where $\overline{d}(x, y) = \min\{1, d(x, y)\}$, is totally bounded.

Theorem 2.17.11 Let $\{(X, d_i)\}_{i=1}^{\infty}$ be a family of metric spaces such that d_i is bounded by 1 for each $i = 1, 2, \ldots$, then the product $X = \prod_{i=1}^{\infty} X_i$ with the product metric D is totally bounded if and only if each (X_i, d_i) is totally bounded.

Proof. Assume that the product metric space (X, D) is totally bounded. Then the subspace $Y_j = \prod_{i=1}^{\infty} A_i$, where $A_j = X_j$, and $A_i = \{x_i\} \subseteq X_i$ for all $i \neq j$, is totally bounded by Theorem 2.17.10. Let $\varepsilon > 0$ be a real number. Then there exists a finite ε/j-net A in Y_j. The j^{th} projection of A, $p_j(A)$, is a finite ε-net in X_j. Thus, X_j is totally bounded. Conversely, assume that the metric spaces (X_j, d_j) are totally bounded. Let $\varepsilon > 0$ be a real number. Let k be a positive integer such that $1/k < \varepsilon/2$. Then for all $i \leq k$, choose a finite $\varepsilon/2$-net $A_i = \{x_1^i, x_2^i, \ldots, x_{m(i)}^i\}$, and for $i > k$, choose $A_i = \{x_0^i\} \subseteq X_i$. Let $A = \prod_{i=1}^{\infty} A_i$. Clearly A is a finite subset of X. We show that A is an ε-net in X. Let $x = (y_i)$ be a point in X. For every $i \leq k$, there exists an $x_{j_i}^i \in A_i$ such that $d_i(x_{ji}^i, y_i) < \varepsilon/2$. Then the point $y = (x_{j_1}^1, x_{j_2}^2, \ldots, x_{j_k}^k, x_0^{k+1}, x_0^{k+2}, \ldots)$ is in A, and

$$\begin{aligned} D(x, y) &= \max\{d_1(x_{j_1}^1, y_1)/1, d_2(x_{j_2}^2, y_2)/2, \ldots, d_k(x_{j_k}^k, y_k)/k, \sup_{i>k}\{d_i(x_0^i, y_i)/i\}\} \\ &\leq \varepsilon/2 < \varepsilon, \end{aligned}$$

so that A is a finite ε-net for X.

EXERCISE

1. Let (X, d_1) and (X, d_2) be metric spaces, and $d(x, y) = d_1(x, y) + d_2(x, y)$ for all x and y in X. Then d is a metric on X, and if (X, d_1) and (X, d_2) are totally bounded, then (X, d) is totally bounded.

2. Show that the closed unit sphere S in the metric space $\mathcal{C}[0, 1]$ with sup metric is bounded but not totally bounded. On the other hand, show that the subset of S given by
$$A_L = \{f \in \mathcal{C}[0, 1] : |f(t) - f(s)| \leq L|t - s|\}, \ L \geq 0,$$
is totally bounded.

3. Show that the Hilbert cube $A = \{\langle x_n \rangle \in l_2 : |x_n| \leq 1/n\}$ in l_2-space is totally bounded.

4. Let $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined by
$$d(x, y) = [(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2]^{1/2},$$
and $A = \left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} |x_i|^2 \leq 1\right\}$.
Show that A is totally bounded in $(\mathbb{R}^n, d)$.

Chapter 3

Sequences in Metric Spaces

3.1 The Limit of a Sequence

The reader must have seen in an elementary course of real analysis that the study of convergence of real sequences is carried out primarily to analyse the convergence of several types of infinite series of real numbers which, in turn, occur as solutions of differential equations occurring in many practical applications. As a further application, the notion of continuity of a function is characterised in terms of sequences, viz., a function f is continuous at x if and only if every sequence $\langle x_n \rangle$ converging to x implies that $\langle f(x_n) \rangle$ converges to $f(x)$. In metric spaces, the notion of convergence of a sequence plays an enhanced role. We have already seen several metric spaces of sequences. The notion of convergence of a sequence in a metric space has been introduced in Section 2.6. To study it in greater detail, we make it explicit in the following definition.

Definition 3.1.1 Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence in X. Then $\langle X_n \rangle$ is said to converge to a point $x \in X$ if for each real number $\varepsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that $d(x_n, x) < \varepsilon$ for all $n \geq m$.

The point x in the above definition is called a limit of the sequence $\langle x_n \rangle$, and we write $\lim_{n \to \infty} x_n = x$ or simply as $x_n \to x$, or as $\langle x_n \rangle \to x$.

Examples 3.1.2

1. Let (X, d) be a discrete metric space, and $\langle x_n \rangle$ be a sequence in X which converges to x in X. Let $\varepsilon = \frac{1}{2}$. Since $\langle x_n \rangle$ converges to x, there exists an $m \in \mathbb{Z}_+$ such that $d(x_n, x) < \frac{1}{2}$ for all $n \geq m$. This means that $x_n = x$ for $n \geq m$, that is, the sequence is of the form $\langle x_1, x_2, \ldots, x_{m-1}, x, x, x, \ldots \rangle$ which has the property that it becomes constant after a certain number of terms. Such a sequence is called an eventually constant sequence. Thus, the convergent sequences in a discrete metric space are eventually constant sequences.

2. In $\mathbb{R}^2$, the sequence $\left\langle \left(\frac{1}{n}, \frac{1}{n}\right) \right\rangle$ is a convergent sequence which converges to (0,0): let $\varepsilon > 0$ be a real number. If we choose $m \in \mathbb{Z}_+$ to be greater than $\frac{\sqrt{2}}{\varepsilon}$, then

$$d\left(\left(\frac{1}{n}, \frac{1}{n}\right), (0,0)\right) = \left[\left(\frac{1}{n} - 0\right)^2 + \left(\frac{1}{n} - 0\right)^2\right]^{\frac{1}{2}} < \varepsilon \text{ for all } n \geq m.$$

3. Let $X =]0, 1]$ be equipped with usual metric $d(x, y) = |x - y|$ for all x, y in X. Then $\left\langle \frac{1}{n} \right\rangle$ is a sequence in X which does not converge to any point in X. Let p be any point in X. Suppose that $\left\langle \frac{1}{n} \right\rangle$ converges to p. Let $\varepsilon = \frac{p}{2}$. Then there exists a positive integer m such that
$$d\left(\frac{1}{n}, p\right) = \left|\frac{1}{n} - p\right| < \frac{p}{2} \text{ for all } n \geq m.$$
That is, $\frac{p}{2} < \frac{1}{n}$ for all $n \geq m$ which is not possible. Thus, $\left\langle \frac{1}{n} \right\rangle$ does not converge to the point p. Since p is any arbitrary point in X, $\left\langle \frac{1}{n} \right\rangle$ does not converge to any point in X.

4. Let $f_n : [0, 1] \to \mathbb{R}$ be the function $f_n(x) = x^n$. Then the sequence $\langle f_n(x) \rangle$ converges in $\mathbb{R}$ for each $x \in [0, 1]$.

An immediate consequence of the above definition is the following equivalence.

Theorem 3.1.3 A sequence $\langle x_n \rangle$ in a metric space (X, d) converges to a point x in X if and only if for each real number $\varepsilon > 0$, there exists a positive integer $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$.

Proof. The proof is straightforward.

Theorem 3.1.4 A sequence $\langle x_n \rangle$ in a metric space (X, d) converges to a point x in X if and only if for each neighbourhood N of x, there exists an $m \in \mathbb{Z}_+$ such that $x_n \in N$ for all $n \geq m$.

Proof. If N is a neighbourhood of x, then there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon) \subseteq N$. If $\langle x_n \rangle$ converges to x, then there exists an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$, and the necessity part follows. Conversely, since for each real number $\varepsilon > 0$, $S(x, \varepsilon)$ is a neighbourhood of x, by our assumption, there exists an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$. Thus, $\langle x_n \rangle$ converges to x.

Theorem 3.1.5 A sequence in a metric space can not converge to more than one point.

Proof. Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence in X which converges to two distinct points x and y in X. Taking $\varepsilon = \frac{1}{2}d(x, y)$, and considering the fact that $\langle x_n \rangle \to x$, there exists an $m_1 \in \mathbb{Z}_+$ such that $d(x_n, x) < \varepsilon$ for all $n \geq m_1$. Again since $\langle x_n \rangle \to y$, there exists an $m_2 \in \mathbb{Z}_+$ such that $d(x_n, y) < \varepsilon$ for all $n \geq m_2$. Then, by the triangle inequality
$$d(x, y) \leq d(x, x_n) + d(x_n, y) < \varepsilon + \varepsilon \text{ for all } n \geq \max\{m_1, m_2\}.$$
Thus, $d(x, y) < 2\varepsilon$. Since ε is arbitrary, $d(x, y) = 0$ which means that $x = y$, a contradiction.

The above theorem proves the uniqueness of a limit, and therefore we speak of "the limit" instead of "a limit" of a convergent sequence.

The following result is very useful.

Lemma 3.1.6 Let (X, d) be a metric space. Then the following inequality holds.
$$|d(x_1, y_1) - d(x_2, y_2)| \leq d(x_1, x_2) + d(y_1, y_2) \text{ for all } x_1, x_2, y_1 \text{ and } y_2 \text{ in } X.$$

Proof. By the triangle inequality,
$$d(x_1, y_1) \leq d(x_1, x_2) + d(x_2, y_2) + d(y_2, y_1).$$
Therefore, $d(x_1, y_1) - d(x_2, y_2) \leq d(x_1, x_2) + d(y_1, y_2)$. Now by interchanging x_1 with x_2, and y_1 with y_2, the desired inequality follows.

Theorem 3.1.7 Let (X, d) be a metric space, and $\langle x_n \rangle$ and $\langle y_n \rangle$ be convergent sequences in X converging to x and $y \in X$, respectively. Then the sequence $\langle d(x_n, y_n) \rangle_{n=1}^{\infty}$ converges in the real line $\mathbb{R}$ to $d(x, y) \in \mathbb{R}$.

Proof. By Lemma 3.1.6, we have

$$|d(x_n, y_n) - d(x, y)| \leq d(x_n, x) + d(y_n, y),$$

and the required convergence immediately follows.

Theorem 3.1.8 Let (X, d) be a metric space. A sequence $\langle x_n \rangle$ in X converges to $x \in X$ if and only if the sequence $\langle d(x_n, x) \rangle$ converges to 0 in $\mathbb{R}$.

Proof. Let $\varepsilon > 0$ be a real number. The sequence $\langle x_n \rangle$ converges in X to x if and only if there exists an $m \in \mathbb{Z}_+$ such that $d(x_n, x) < \varepsilon$ for all $n \geq m$ if and only if $|d(x_n, x) - 0| < \varepsilon$ for all $n \geq m$ if and only if the sequence $\langle d(x_n, x) \rangle$ converges in $\mathbb{R}$ to $0 \in \mathbb{R}$.

A double sequence in a set X is a mapping $s : N \times N \to X$. If $s(n, m) = x_{n,m}$, then the double sequence s may be denoted in terms of its values as $\langle (x_{n,m} \rangle_{n,m=1}^{\infty}$. A double sequence $\langle x_{n,m} \rangle_{n,m=1}^{\infty}$ or simply $\langle x_{n,m} \rangle$ in a metric space (X, d) is said to converge to a point x in X if for each real number $\varepsilon > 0$, there exists a positive integer p such that

$$d(x_{n,m}, x) < \varepsilon \text{ for all } n, m \geq p.$$

Now, it may be noted that if $\langle x_n \rangle$ is a convergent sequence in a metric space (X, d) converging to $x \in X$, then the double sequence $\langle d(x_n, x_m) \rangle$ in $\mathbb{R}$ converges to zero since $0 \leq d(x_n, x_m) \leq d(x_n, x) + d(x, x_m)$.

Recall the definition of a subsequence of a sequence, and we have the following.

Theorem 3.1.9 Let (X, d) be a metric space. A sequence $\langle x_n \rangle$ in X converges to a point $x \in X$ if and only if every subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ converges to x.

Proof. Let $\varepsilon > 0$ be a real number. If $\langle x_n \rangle \to x$, then there exists an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$. This means that $x_{n_k} \in S(x, \varepsilon)$ for all $k \geq m$ since $n_k \geq k$ for all $k \in \mathbb{Z}_+$. The converse is obvious.

The sufficiency condition for the convergence of a sequence in the above theorem can be replaced as in the following theorem.

Theorem 3.1.10 Let (X, d) be a metric space. A sequence $\langle x_n \rangle$ in X converges to $x \in X$ if and only if every subsequence of $\langle x_n \rangle$ has a subsequence that converges to x.

Proof. Since a subsequence of a subsequence of a sequence $\langle x_n \rangle$ is a subsequence of $\langle x_n \rangle$, the necessity part follows by the previous theorem. Conversely, assume that every subsequence of $\langle x_n \rangle$ has a subsequence that converges to $x \in X$. Now suppose, on the contrary, that $\langle x_n \rangle$ does not converge to x. So there exists a real number $\varepsilon > 0$ such that for each $m \in \mathbb{Z}_+$, there is a positive integer $n > m$ such that $x_n \notin S(x, \varepsilon)$. But then we get a subsequence of $\langle x_n \rangle$ which has no subsequence that converges to x.

Example 3.1.11 Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence in X that converges to $x \in X$. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be any permutation (a bijective mapping) of $\mathbb{N}$. Then the sequence $\langle x_{\sigma(n)} \rangle_{n=1}^{\infty}$ also converges to x. To see this, let $\varepsilon > 0$ be a real numbers. Then there exists an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$. Now let m_0 be the largest positive integer such that $x_{\sigma(m_0)} \in \{x_1, x_2, \ldots, x_{m-1})\}$. Such an integer exists since the points $x_1, x_2, \ldots, x_{m-1}$ are finite in number. Now it follows that $x_{\sigma(n)} \in S(x, \varepsilon)$ for all $n \geq m_0$.

EXERCISE

1. Let $\langle x_n \rangle$ be a convergent sequence in a metric space (X, d), and i be a fixed positive integer. Show that the sequence $\langle d(x_i, x_n) \rangle_{n=1}^{\infty}$ is a bounded sequence in the real line $\mathbb{R}$.

2. Let $\langle x_n \rangle$ be a sequence in the real line $\mathbb{R}$ such that $x_n \geq 0$, and $\lim_{n \to \infty} x_n = 0$. Show that there are infinitely many indices n such that $x_m \leq x_n$ for all $m \geq n$.

3. Let $\langle x_n \rangle$ be a sequence in a metric space. Show that if the three subsequences $\langle x_{2n} \rangle$, $x_{2n+1} \rangle$ and $\langle x_{3n} \rangle$ are convergent, $\langle x_n \rangle$ is convergent.

4. Give an example of sequence $\langle x_n \rangle$ in $\mathbb{R}$ which is not convergent but for each $k \geq 2$, the subsequence $\langle x_{kn} \rangle_{n=1}^{\infty}$ is convergent. (Consider the subsequence $\langle x_{n_k} \rangle$, where $\langle n_k \rangle$ is the strictly increasing sequence of prime numbers.)

5. If $\langle z_n \rangle$ is a sequence in the complex plane $\mathbb{C}$ converging to $z \in \mathbb{C}$, prove that $\langle |z_n| \rangle$ converges to $|z|$ in $\mathbb{R}$. Hence, show that the subset $\{w \in \mathbb{C} : |w| \leq c\}$ of $\mathbb{C}$ is a closed set for any positive real number c.

6. Let Y be the set of all complex valued sequences $\langle z_n \rangle$ for which $|z_n| \leq \frac{1}{n}$ for all $n \in \mathbb{Z}_+$. Define d by

   $d(x, y) = \left[\sum_{n=1}^{\infty} |z_n - z'_n|^2 \right]^{\frac{1}{2}}$ for all $x = \langle z_n \rangle$ and $y = \langle z'_n \rangle$ in Y. Show that (Y, d) is a subspace of l_2, and that Y is a closed set.

7. Let (X, d) be the pseudometric space of Exercise 1, Section 2.14 with $a = 0, b = \frac{1}{2}$. Show that the sequence $\langle x_n \rangle$, where $x_n(t) = t^n$, $0 \leq t \leq \frac{1}{2}$ is convergent, and that any constant function on $\left[0, \frac{1}{2}\right]$ serves as its limit. Hence convergent sequence in a pseudometric space need not have a unique limit.

8. Let A and B be subsets of m dimensional Euclidean space $(\mathbb{R}^m, d)$, and A be bounded. Show that there exist points $a \in A$ and $b \in B$ such that $d(a, b) = d(A, B)$. (Choose $a_n \in A, b_n \in B$, such that $d(a_n, b_n)$ converges to $d(A, B)$, and consider convergent subsequence of $\langle a_n \rangle$.)

9. Consider the sequence $s = \left\langle \frac{1}{k} \right\rangle_{k=1}^{\infty}$. Find a sequence $\langle x_n \rangle$ in l_2-space such that each x_n is distinct from s, and such that $\langle x_n \rangle$ converges to s in l_2-space.

3.2 Cluster Point of a Sequence

Definition 3.2.1 Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence in X. A point x in X is said to be a cluster point (or limit point) of $\langle x_n \rangle$ if for each real number $\varepsilon > 0$ and each $m \in \mathbb{Z}_+$, there exists an $n \in \mathbb{Z}_+$ such that $n > m$, and $d(x_n, x) < \varepsilon$.

Example 3.2.2 Let (X, d) be a metric space, then the sequence $\langle x, y, x, y, x, y, x, \ldots \rangle$ in which x and y alternate, has two cluster points x and y.

The following theorems are immediate consequences of the above definition of a cluster point of a sequence.

Theorem 3.2.3 Let (X, d) be a metric space. A point x in X is a cluster point of a sequence $\langle x_n \rangle$ in X if and only if for each real number $\varepsilon > 0$, and for each $m \in \mathbb{Z}_+$, there is an n in $\mathbb{Z}_+$ such that $n > m$, and $x_n \in S(x, \varepsilon)$.

Theorem 3.2.4 Let (X, d) be a metric space. A point x in X is a cluster point of a sequence $\langle x_n \rangle$ in X if and only if for each neighbourhood N of x, and for each $m \in \mathbb{Z}_+$, there is an $n \in \mathbb{Z}_+$ such that $n > m$, and $x_n \in N$.

In view of the above theorem, sometimes, we say that a point x is a cluster point of a sequence if the points of the sequence are frequently in every neighbourhood of x.

Theorem 3.2.5 The limit of a convergent sequence is a cluster point of the sequence.

Proof. Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence in X converging to x in X. Then for each real number $\varepsilon > 0$, there is an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$. Now for each $p \in \mathbb{Z}_+$, select $n > \max\{p, m\}$. Then $x_n \in S(x, \varepsilon)$, and so x is a cluster point of the sequence.

Theorem 3.2.6 A convergent sequence in a metric space has a unique cluster point.

Proof. Let (X, d) be a metric space, and $\langle x_n \rangle$ be a sequence converging to $x \in X$. By Theorem 3.2.5, x is a cluster point of $\langle x_n \rangle$. Let y be another cluster point of $\langle x_n \rangle$ different from x. Let $\varepsilon = \frac{1}{2} d(x, y)$. Since x is the limit of $\langle x_n \rangle$, there exists an $m \in \mathbb{Z}_+$ such that $d(x_n, x) < \varepsilon$ for $n \geq m$. Since y is a cluster point, there is an n in $\mathbb{Z}_+$ such that $n > m$, and $d(x_n, y) < \varepsilon$. Then

$$d(x, y) < d(x, x_n) + d(x_n, y) < 2\varepsilon$$

which is a contradiction.

The converse of the above theorem is not true, that is, if a sequence has a unique cluster point, then it may not be true that the sequence is convergent. Consider the sequence $\langle x_n \rangle$ in $\mathbb{R}^2$ defined by

$$x_n = \begin{cases} \left(\frac{1}{n}, 0\right) & \text{if } n \text{ is even,} \\ (1, 1) & \text{if } n \text{ is odd.} \end{cases}$$

It is easily seen that (0,0) is a cluster point of $\langle x_n \rangle$, but $\langle x_n \rangle$ does not converge to any point in $\mathbb{R}^2$.

From the definition of a bounded sequence of real numbers, it is immediate that a sequence $\langle x_n \rangle$ of real numbers is bounded if and only if the set of points of the sequence $\{x_n : n \in \mathbb{Z}_+\}$ is a bounded set. This result is made the definition of a bounded sequence in metric spaces.

Definition 3.2.7 Let (X, d) be a metric space. A sequence $\langle x_n \rangle$ in X is said to be bounded if the set of points of the sequence $\{x_n : n \in \mathbb{Z}_+\}$ is a bounded set.

Examples 3.2.8

1. In $\mathbb{R}^n$, for each $m \in Z_+$, let y_m be a point in the set $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : [x_1^2 + x_2^2 + \ldots + x_n^2]^{\frac{1}{2}} < 1\}$. Then the sequence $\langle y_m \rangle$ is a bounded sequence in $\mathbb{R}^n$.
2. In discrete metric space, every sequence is a bounded sequence.
3. If a metric space (X, d) has the bounded metric d, then every sequence in it is a bounded sequence.

Theorem 3.2.9 (Bolzano-Weierstrass) Every bounded sequence in $\mathbb{R}^n$ has a cluster point.

Proof. Let $\langle x_k \rangle$ be a bounded sequence in $\mathbb{R}^n$, and A be the set of points of this sequence, that is, $A = \{x_k : k \in \mathbb{Z}_+\}$. If A is finite, then for some point x in $\mathbb{R}^n$, $x_k = x$ for infinitely many values of k. Consequently, if $\varepsilon > 0$ is a real number, and $p \in \mathbb{Z}_+$, then there exists an $m > p$ such that $x_m = x$, and so $d(x_m, x) = 0 < \varepsilon$, where d is the Euclidean metric on $\mathbb{R}^n$. Thus, x is a cluster point in this case. So we assume that A is infinite. By Corollary 2.8.14, A has a limit point y in $\mathbb{R}^n$. Therefore, for each real number $\varepsilon > 0$, the neighbourhood $S(x, \varepsilon)$ of x contains infinitely many points of A. Hence, for each $p \in \mathbb{Z}_+$ there exists an $m \in \mathbb{Z}_+$ such that $m > p$, and $x_m \in S(x, \varepsilon)$. Thus, x is a cluster point of $\langle x_k \rangle$.

In general, a bounded sequence in a metric space may not have any cluster point. For example, let (X, d) be an infinite discrete metric space. Then any sequence $\langle x_n \rangle$ in X such that $x_i \neq x_j$ for $i \neq j$ is a bounded sequence which has no cluster point.

Theorem 3.2.10 Let (X, d) be a metric space. A point x in X is a cluster point of a sequence $\langle x_n \rangle$ in X if and only if there exists a subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ that converges to x.

Proof. First assume that x is a cluster point of $\langle x_n \rangle$. Then there exists an $n_1 \in \mathbb{Z}_+$ such that $d(x_{n_1}, x) < 1$. Again, there exists n_2 such that $n_2 > n_1$, and $d(x_{n_2}, x) < \frac{1}{2}$. By induction, there exists $n_k \in \mathbb{Z}_+$ such that $n_k > n_{k-1}$, and $d(x_{n_k}, x) < \frac{1}{k}$. Thus, $\langle x_{n_k} \rangle$ is a subsequence of $\langle x_n \rangle$ which converges to x. Conversely, assume that the subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ converges to $x \in X$. Let $\varepsilon > 0$ be a real number, and $m \in \mathbb{Z}_+$. Since $\langle x_{n_k} \rangle$ converges to x, there exists a $p \in \mathbb{Z}_+$ such that $d(x_{n_k}, x) < \varepsilon$ for all $k \geq p$. Then for $r > \max\{p, m\}$, $n_r > m$, and $d(x_{n_r}, x) < \varepsilon$. Thus, x is a cluster point of $\langle x_n \rangle$.

Corollary 3.2.11 Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Theorem 3.2.12 Let (X, d) be a metric space. A point $x \in X$ is an adherent point of a subset A of X if and only if there is a sequence in A converging to x.

Proof. Let x be an adherent point of A. If x is in A, then the constant sequence $\langle x, x, \ldots \rangle$ is in A that converges to x. If x is not in A, then x is a limit point of A. Therefore, for each $n \in \mathbb{Z}_+$, the open sphere $S\left(x, \frac{1}{n}\right)$ intersects with A, and so there is a point x_n in $S\left(x, \frac{1}{n}\right) \cap A$. Then $\langle x_n \rangle$ is a sequence in A which converges to x. Conversely, assume that $\langle x_n \rangle$ is a sequence in A that converges to x. Then for each real number $\varepsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq m$. Thus, x is an adherent point of A.

Corollary 3.2.13 A set A in a metric space (X, d) is closed if and only if the limit of every convergent sequence in A is in A.

Proof. Apply Theorem 3.2.12.

EXERCISE

1. Prove that the set of all cluster points of a sequence in a metric space is a closed set. Also show that if X is a separable metric space, and F is a closed subset of X, then there exists a sequence $\langle x_n \rangle$ in X such that the set of all cluster points of $\langle x_n \rangle$ is equal to F.

2. Show that if A is a countable set, then there exists a sequence $\langle x_n \rangle$ in A such that for

each element x in A, $x_n = x$ for infinitely many values of n, conclude from this that if A is any countable subset of a metric space X, then there exists a sequence $\langle x_n \rangle$ in X such that each point of A is a cluster point of $\langle x_n \rangle$.

3.3 Convergence in Products

Let $(X_i, d_i), i = 1, 2, \ldots, k$ be metric spaces. Let $\langle y_n \rangle$ be a sequence in $X = \prod_{i=1}^{k} X_i$, where $y_n = (x_{n1}, x_{n2}, \ldots, x_{nk})$ for each $n \in \mathbb{Z}_+$. Assume that $y = (x_1, x_2, \ldots, x_k)$ is a point in X. Let d be a metric on X. If $\langle y_n \rangle$ converges to y, then y is sometimes, called the pointwise limit. The sequence $\langle x_{nj} \rangle_{n=1}^{\infty}$ is the sequence of j^{th} co-ordinates of y_n's. When $\langle y_n \rangle$ converges to y, it may or may not be true that the sequence $\langle x_{nj} \rangle_{n=1}^{\infty}$ converges in X_j to j^{th} co-ordinate of y for each j. If the statement : $\langle y_n \rangle$ converges to y if and only if $\langle x_{nj} \rangle_{n=1}^{\infty}$ converges to x_j for each $j = 1, 2, \ldots, k$, is true, we say in a bit loose language that pointwise limit is equivalent to co-ordinate wise (or termwise) limits, where, infact, the pointwise limit is made up of co-ordinate wise limits. Any metric d on X having this property viz., the pointwise limit is equivalent to the co-ordinate wise limit for a sequence in X, is called a product metric.

The metric introduced in Definition 2.15.2, infact, has the above property, and this justifies its name as a product metric. With the notations as in the above paragraph, the following theorem states precisely this for the product metric d on X.

Theorem 3.3.1 A sequence $\langle y_n \rangle$ in X converges to y if and only if the sequence $\langle x_{nj} \rangle_{n=1}^{\infty}$ in the j^{th} co-ordinates converges in X_j to x_j for each $j = 1, 2, \ldots, k$.

Proof. If $\langle y_n \rangle$ converges to y, then for each real number $\varepsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that $d(y_n, y) < \varepsilon$ for all $n \geq m$. This means that $\max_j \{d_j(x_{nj}, x_j)\} < \varepsilon$ for all $n \geq m$. Therefore, $d_j(x_{nj}, x_j) < \varepsilon$ for all $n \geq m$ that is, $\langle x_{n_j} \rangle_{n=1}^{\infty}$ is convergent, and converges to x_j for each j. Conversely, let $\varepsilon > 0$ be a real number. By our assumption, for each j, there exists an $m_j \in \mathbb{Z}_+$ such that $d_j(x_{nj}, x_j) < \varepsilon$ for all $n \geq m_j$. Taking m_0 to be the maximum of $m_1, m_2, \ldots, m_k$, it follows that $d(y_n, y) < \varepsilon$ for all $n \geq m_0$ and hence, $\langle y_n \rangle$ converges to y.

Interestingly, there are other metrics on the product X which behave like the product metric.

Theorem 3.3.2 Let $(X_i, d_i), i = 1, 2, \ldots, k$ be metric spaces, and $X = \prod_{i=1}^{k} X_i$. Then for each real number $p \geq 1$, the mapping ρ_p defined by

$$\rho_p(x, y) = \left[\sum_{i=1}^{k} \{d_i(x_i, y_i)\}^p \right]^{\frac{1}{p}}$$

for all $x = (x_1, x_2, \ldots, x_k)$, $y = (y_1, y_2, \ldots, y_k)$ in X is a product metric on X.

Proof. That ρ_p is a metric on X is easy to prove. The triangle inequality for ρ_p is established

as follows using Minkowski's inequality (Lemma 2.1.6):

$$\left[\sum_{i=1}^{k}\{d_i(x_i,y_i)\}^p\right]^{\frac{1}{p}} \leq \left[\sum_{i=1}^{k}\{d_i(x_i,z_i)+d_i(z_i,y_i)\}^p\right]^{\frac{1}{p}}$$
$$\leq \left[\sum_{i=1}^{k}\{d_i(x_i,z_i)\}^p\right]^{\frac{1}{p}} + \left[\sum_{i=1}^{k}\{d_i(z_i,y_i)\}^p\right]^{\frac{1}{p}}.$$

To prove that d is a product metric, let $\langle y_n\rangle$, where $y_n = (x_{n1}, x_{n2}, \ldots, x_{nk})$ be a sequence in X which converges to $y = (x_1, x_2, \ldots, x_k)$ in X. Since $d_i(x_{ni}, x_i) \leq \rho_p(y_n, y)$, the convergence of the sequence $\langle x_{ni}\rangle_{n=1}^{\infty}$ in the i^{th} co-ordinates to x_i is immediate for each $i = 1, 2, \ldots, k$. Conversely, assume that each co-ordinate sequence $\langle x_{ni}\rangle_{n=1}^{\infty},\ i = 1, 2, \ldots, k$ converges to x_i. Let $\varepsilon > 0$ be a real number. Then there exists an $m_i \in \mathbb{Z}_+$ such that

$$d_i(x_{ni}, x_i) < \left(\frac{\varepsilon^p}{k}\right)^{\frac{1}{p}} \text{ for all } n \geq m_i.$$

Taking $m_0 = \max\{m_1, m_2, \ldots, m_k\}$, it follows that

$$\left[\sum_{i=1}^{k}\{d_i(x_{ni}, x_i)\}^p\right]^{\frac{1}{p}} < \varepsilon \text{ for all } n \geq m_0.$$

It may be noted that the metric d_p in Example 2.1.8 is a product metric of n copies of $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

In Example 2.1.22, we introduced the real (complex) Fréchet space. This can be generalized as follows : Let $(X_i, d_i), i \in \omega$ be metric spaces, and $X = \prod_{i\in\omega} X_i$. Then the mapping σ defined by

$$\sigma(x, y) = \sum_{i=1}^{\infty} \frac{d_i(x_i, y_i)}{2^i(1 + d_i(x_i, y_i))}$$

for all $x = \langle x_n\rangle$ and $y = \langle y_n\rangle$ in X is a metric on X which is also a product metric. The metric space (X, σ) is known as a general Fréchet space.

Theorem 3.3.3 The metric σ in the general Fréchet metric space (X, σ) is a product metric.

Proof. Let $\langle y_n\rangle$, where $y_n = \langle x_{n1}, x_{n2}, \ldots\rangle$, be a sequence in X converging to $y = \langle x_1, x_2, \ldots\rangle$. Since

$$0 \leq \frac{d_j(x_{nj}, x_j)}{2^j(1 + d_j(x_{nj}, x_j))} \leq \sigma(y_n, y),$$

it follows that

$$\lim_{n\to\infty} \frac{d_j(x_{nj}, x_j)}{2^j(1 + d_j(x_{nj}, x_j))} = 0.$$

Since the denominator in $\frac{d_j(x_{nj},x_j)}{2^j(1+d_j(x_{nj},x_j)}$, is at least one,

$$\lim_{n\to\infty} d_j(x_{nj}, x_j) = 0,$$

that is, the sequence $\langle x_{nj}\rangle_{n=1}^{\infty}$ in the j^{th} co-ordinate of $\langle y_n\rangle$ converges to x_j for each $j \in \omega$.

Conversely, assume that $\langle x_{nj}\rangle_{n=1}^{\infty}$ converges to x_j for each $j \in \omega$. Let $\varepsilon > 0$ be a real number. Then there exists an $n_0 \in \mathbb{Z}_+$ depending on ε such that $\sum_{i=n_0+1}^{\infty} \frac{1}{2^i} < \frac{\varepsilon}{2}$. Also since for each $i = 1, 2, \ldots, n_0$, the sequence $\langle x_{ni}\rangle_{n=1}^{\infty}$ converges to x_i, there exists a $k_i \in \mathbb{Z}_+$ such that

$$d_i(x_{ni}, x_i) < \frac{\varepsilon}{2n_0} \text{ for all } n \geq k_i.$$

Assume that $k_0 = \max\{k_1, \ldots, k_{n_0}\}$. Then

$$\begin{aligned}\sigma(y_n, y) &\leq \sum_{i=1}^{n_0} d_i(x_{ni}, x_i) + \sum_{i=n_0+1}^{\infty} \frac{1}{2^i} \\ &< \frac{n_0\varepsilon}{2n_0} + \frac{\varepsilon}{2} \text{ for } n \geq k_0.\end{aligned}$$

Thus, $\sigma(y_n, y) < \varepsilon$ for all $n \geq k_0$.

The following theorem considers the metric space l_p of Example 2.1.20.

Theorem 3.3.4 Let $\langle y_n\rangle$ be a sequence in l_p-space, where $y_n = \langle x_{n1}, x_{n2}, \ldots\rangle$, converging to $y = \langle x_1, x_2, \ldots\rangle$ in l_p. Then the sequence $\langle x_{nj}\rangle_{n=1}^{\infty}$ in the j^{th} co-ordinate of $\langle y_n\rangle$ converges to x_j for each $j \in \mathbb{Z}_+$.

Proof. Since

$$0 \leq |x_{nj} - x_j| \leq \left[\sum_{i=1}^{\infty} |x_{ni} - x_i|^p\right]^{\frac{1}{p}} = d_p(y_n, y),$$

the conclusion follows.

The converse of the above theorem is not true: let us define a sequence $\langle y_n\rangle$ in l_p-space as follows. Assume that $y_n = \langle x_{n1}, x_{n2}, \ldots\rangle$, where

$$x_{nj} = \begin{cases} 1 & \text{if } n = j, \\ 0 & \text{if } n \neq j. \end{cases}$$

Then $\lim_{n\to\infty} x_{nj} = 0$ for each $j \in \mathbb{Z}_+$, but the sequence $\langle y_n\rangle$ does not converge to $0 = \langle 0, 0 \ldots\rangle$ in l_2-space since $d_p(y_n, 0) = 1$ for all n.

The metric D in Theorem 2.15.10 was also called product metric. Here is a justification for that claim with the notation borrowed.

Theorem 3.3.5 The metric D is a product metric in the metric space (X, D).

Proof. Let $\langle y_n\rangle$, where $y_n = \langle x_{n1}, x_{n2} \ldots\rangle$ be a sequence in X converging to $y = \langle x_1, x_2, \ldots\rangle$ in X. Then the necessity part follows from the inequality

$$0 \leq \frac{\rho_i(x_{ni}, x_i)}{i} \leq \text{lub}_{i\in\omega} \left\{\frac{\rho_i(x_{ni}, x_i)}{i}\right\} = D(y_n, y)$$

and the equivalence of ρ_i and d_i. Conversely, assume that $\langle x_{ni}\rangle_{n=1}^{\infty}$ converges to x_i in (X_i, d_i) for each $i \in \omega$. Then $\langle x_{ni}\rangle_{n=1}^{\infty}$ converges to x_i in (X_i, ρ_i) for each $i \in \omega$. Also there exists an $n_0 \in \mathbb{Z}_+$ such that $\frac{1}{i} < \frac{\varepsilon}{2}$ for all $i \geq n_0$. Also since for each $i = 1, 2, \ldots, n_0$, the sequence $\langle x_{ni}\rangle_{n=1}^{\infty}$ converges to x_i, there exists a $k_i \in \mathbb{Z}_+$ such that

$$\rho_i(x_{ni}, x_i) < \frac{\varepsilon i}{2} \text{ for all } n \geq k_i.$$

Assuming $k_0 = \max\{k_1, k_2, \ldots, k_{n_0}\}$, the following inequality holds:

$$D(y_n, y) \leq \max_{1 \leq i \leq n_0} \left\{ \frac{\rho_i(x_{ni}, x_i)}{i} \right\} + \text{lub}_{i > n_0} \left\{ \frac{1}{i} \right\}$$
$$< \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon \text{ for all } n \geq k_0.$$

Thus, the sequence $\langle y_n \rangle$ converges to y.

3.4 Characterisation of Equivalent Metrics

Theorem 3.4.1 Let (X, d_1) and (X, d_2) be metric spaces. The metrics d_1 and d_2 are equivalent if and only if every subset of X has the same closure in (X, d_1) and (X, d_2).

Proof. Let d_1 and d_2 be equivalent metrics on X, and A be a subset of X. Let $x \in X - A$ be a limit point of A in (X, d_2). Then there exists a sequence $\langle x_n \rangle$ in A converging to x in (X, d_2). By Theorem 2.6.7, the sequence $\langle x_n \rangle$ converges to x in (X, d_1). Thus, x is a limit point of A in (X, d_1). Hence, the closure of a set A in (X, d_1) and (X, d_2) are equal. Conversely, suppose that d and d' are not equivalent. Also, without loss of generality, assume that $\langle x_n \rangle$ is a sequence in X which converges to a point x in (X, d_1), and $\langle x_n \rangle$ does not converge in (X, d_2) to x. Then there exists a real number $\varepsilon > 0$ such that for each $m \in \mathbb{Z}_+$, there is a $n > m$ such that $d_2(x_n, x) \geq \varepsilon$. Consequently, there exists a subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ such that $x_{n_k} \notin S(x, \varepsilon, d_2)$ for all $k \in \mathbb{Z}_+$. That is, the subsequence $\langle x_{n_k} \rangle$ is contained in $X - S(x, \varepsilon, d_2)$. Now $X - S(x, \varepsilon, d_2)$ is a closed set in (X, d_2). Let A be the set of points of the subsequence $\langle x_{n_k} \rangle$. Then $A \subseteq X - S(x, \varepsilon, d_2)$ so that the closure of A in (X, d_2) is contained in $X - S(x, \varepsilon, d_2)$. In particular, the closure of A in (X, d_2) does not contain the point x. But the point x is in the closure of A in (X, d_1) since $\langle x_{n_k} \rangle$ is a subsequence of the sequence $\langle x_n \rangle$ converging to x, and therefore $\langle x_{n_k} \rangle$ is itself convergent to x in (X, d_1). Thus, the closure of A in (X, d_1) and (X, d_2) are not identical.

Corollary 3.4.2 Let (X, d_1) and (X, d_2) be metric spaces. If d_1 and d_2 are equivalent, then precisely the same subsets of X are dense in (X, d_1) and (X, d_2).

EXERCISE

1. Let (X, d) be a metric space. Let $\mathcal{H}$ be the set of all closed and bounded subsets of X with Hausdroff metric σ. Let $\langle A_n \rangle$ be a sequence in H which converges to $A \in \mathcal{H}$.
   (a) If $x \in A$, then prove that there is a sequence $\langle x_n \rangle$ in X such that $x_n \in A_n$, and $\langle x_n \rangle$ converges to x.
   (b) If $\langle x_n \rangle$ is a convergent sequence in X such that $x_n \in A_n$, then its limit is in A.

Chapter 4

Limits and Continuity

4.1 Limit and Continuity of Mappings

Recall the notions of limit and continuity of real valued function of real variable from the elementary calculus: let D be a subset of $\mathbb{R}$, and a be a point in $\bar{D}$. Then a function $f : D \to \mathbb{R}$ is said to tend to a real number l (or l is a limit of f) as x in D tends to a if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$|x - a| < \delta \text{ implies that } |f(x) - l| < \epsilon.$$

Since the point a in the above definition need not be a point of D, a function may have limit at a point even if it is not defined at that point.

For the notion of continuity of a real valued function of real variable, the two modifications are required: (i) the point $a \in D$, and (ii) the number l is required, to be equal to $f(a)$. Formally: let $D \subseteq \mathbb{R}$, and $f : D \to \mathbb{R}$ be a function. Then f is said to be continuous at a point $a \in D$ if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$|x - a| < \delta \text{ implies that } |f(x) - f(a)| < \epsilon.$$

In the above definition of limit, the quantity $|x - a|$ is the distance between the point x and the point a, a measure of closeness between x and a, and similarly, $|f(x) - l|$ is the measure of closeness between $f(x)$ and l. Intutively, then the notion of limit says that $f(x)$ comes arbitrarily close to l whenever x comes arbitrarily close to a. The notion of continuity is similarly explained.

In this section, these notions of limit and continuity are extended to arbitrary metric spaces.

Definition 4.1.1 Let (X, d_1) and (Y, d_2) be two metric spaces. Let $A \subseteq X$, and x_0 be any point in $\bar{A}$. Let $f : A \to Y$ be a mapping. Then $f(x)$ is said to tend to a point $y_0 \in Y$ as x in A tends to x_0 if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \text{ and } x \in A \text{ implies that } d_2(f(x), y_0), < \epsilon.$$

This is clearly equivalent to the statement:

$$x \in A \cap S(x_0, \delta) \text{ implies that } f(x) \in S(y_0, \epsilon).$$

When $f(x)$ tends to a point $y_0 \in Y$ as x in A tends to $x_0 \in A$, we write $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y_0$, and say that y_0 is a limit of $f(x)$ at x_0; when $A = X$, we simply write $\lim_{x\to x_0} f(x) = y_0$.

Definition 4.1.2 Let (X, d_1) and (Y, d_2) be two metric spaces. A mapping $f : X \to Y$ is said to be continuous at a point $x_0 \in X$ if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \text{ implies that } d(f(x), f(x_0)) < \epsilon$$

or equivalently

$$x \in S(x_0, \delta) \text{ implies that } f(x) \in S(f(x_0), \epsilon).$$

The following result is an immediate consequence of the above definitions.

Theorem 4.1.3 Let (X, d_1) and (Y, d_2) be two metric spaces, and $f : X \to Y$ be a mapping. Then f is continuous at x_0 if and only if $\lim_{x\to x_0} f(x) = f(x_0)$.

The following theorem proves that a mapping cannot have more than one limit at a point.

Theorem 4.1.4 Let (X, d_1) and (Y, d_2) be metric spaces. Let $A \subseteq X$, and $x_0 \in \overline{A}$. Then a mapping $f : X \to Y$ has a unique limit at x_0, that is, if $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y$, and $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y'$, then $y = y'$.

Proof. Suppose that $y \neq y'$. Let $\epsilon = \frac{d_2(y,y')}{2}$. Since $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y$, there exists a real number $\delta_1 > 0$ such that

$$d_1(x, x_0) < \delta_1 \text{ implies that } d_2(f(x), y) < \epsilon.$$

Again since $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y'$, there exists a real number $\delta_2 > 0$ such that

$$d_1(x, x_0) < \delta_2 \text{ implies that } d_2(f(x), y') < \epsilon.$$

Now assume that $\delta = \min\{\delta_1, \delta_2\}$. Then

$$\begin{aligned} d_2(y, y') &\leq d_2(y, f(x)) + d_2(f(x), y) \\ &< \epsilon + \epsilon \text{ whenever } d_1(x, x_0) < \delta. \end{aligned}$$

Therefore, $d_2(y, y') < 2\epsilon = d_2(y, y')$ which is a contradiction.

The following theorem characterises the notion of limit of a mapping in terms of sequences.

Theorem 4.1.5 Let (X, d_1) and (Y, d_2) be metric spaces, and $A \subseteq X$. Let $x_0 \in \overline{A}$, and $f : A \to Y$ be a mapping. Then $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y_0 \in Y$ if and only if for each sequence $\langle x_n\rangle$ in A converging in X to x_0, the sequence $\langle f(x_n)\rangle$ converges in Y to y_0 if and only if $y_0 \in \overline{f(A)}$ if and only if $y_0 = \lim_{\substack{x\to x_0\\ x\in B}} f(x)$, where $B \subseteq A$, and $x_0 \in \overline{B}$.

Proof. Assume that $\lim_{\substack{x\to x_0\\ x\in A}} f(x) = y_0$. Also assume that $\langle x_n\rangle$is a sequence in A which converges to x_0 in X. Let $\epsilon > 0$ be a real number. Then there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \quad \text{implies that } d_2(f(x), y_0) < \epsilon.$$

Since $\langle x_n \rangle$ converges to x_0 in X, there exists an $m \in \mathbb{Z}_+$ such that

$$d_1(x_n, x_0) < \delta \text{ for all } n \geq m.$$

Therefore, $d_2(f(x_n), y_0) < \epsilon$ for all $n \geq m$, and the necessity part follows. Conversely, suppose that for each sequence $\langle x_n \rangle$ in A converging to x_0 in X, the sequence $\langle f(x_n) \rangle$ converges to y_0 in Y. Suppose on the contrary that $\lim_{\substack{x \to x_0 \\ x \in A}} f(x) \neq y_0$. Then there exists a real number $\epsilon > 0$ such that for each real number $\delta > 0$, there is a point x in A such that

$$d_1(x, x_0) < \delta \text{ but } d_2(f(x), y_0) \not< \epsilon.$$

In particular, for each $n \in \mathbb{Z}_+$, there exists a point x_n in A such that

$$d_1(x_n, x_0) < \frac{1}{n} \quad \text{but } d_2(f(x_n), y_0) \not< \epsilon.$$

Then $\langle x_n \rangle$ is a sequence in A which converges to x_0 in X, but the sequence $\langle f(x_n) \rangle$ does not converges to y_0 in Y. This contradicts our assumption, and the proof of the first equivalence is complete. The proofs of the other equivalences are left to the reader.

The equivalent conditions of continuity of a mapping at a point in the following forms are frequently used.

Theorem 4.1.6 Let (X, d_1) and (Y, d_2) be two metric spaces, and $f : X \to Y$ be a mapping. Let x_0 be a point in X. The following statements are equivalent.

1. f is continuous at x_0.
2. For each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that
$$f(S(x_0, \delta)) \subseteq S(f(x_0), \epsilon).$$
3. For each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that
$$S(x_0, \delta) \subseteq f^{-1}(S(f(x_0)\epsilon)).$$
4. For each neighbourhood N of $f(x_0)$, there exists a neighbourhood M of x_0 such that $f(M) \subseteq N$. (Without loss of generality, the word "neighbourhood" in (4) can be replaced by "open neighbourhood").

Proof. (1) implies (2). By the definition of continuity of f at x_0, for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \text{ implies that } d_2(f(x), f(x_0)) < \epsilon.$$

That is, $x \in S(x_0, \delta)$ implies that $f(x) \in S(f(x_0), \epsilon)$, and so $f(S(x_0, \delta)) \subseteq S(f(x_0), \epsilon)$.

(2) implies (3). By our assumption

$$S(x_0, \delta) \subseteq f^{-1}(f(S(x_0, \delta))) \subseteq f^{-1}(S(f(x_0), \epsilon)).$$

(3) implies (4). Let N be a neighbourhood of $f(x_0)$ in Y. Then there exists a real number $\epsilon > 0$ such that $S(f(x_0), \epsilon) \subseteq N$. Now by our assumption in (3), there exists a real number $\delta > 0$ such that $S(x_0, \delta) \subseteq f^{-1}(S(f(x_0), \epsilon))$. Then

$$f(S(x_0, \delta)) \subseteq f\{f^{-1}(S(f(x_0), \epsilon))\} \subseteq S(f(x_0), \epsilon).$$

and take $M = S(x_0, \delta)$.

(4) implies (1). Let $\epsilon > 0$ be a real number. Then the open sphere $S(f(x_0), \epsilon)$ is a neighbourhood of $f(x_0)$. By our assumption, there is a neighbourhood M of x_0 such that $f(M) \subseteq S(f(x_0), \epsilon)$. Since M is a neighbourhood of x_0, there exists a real number $\delta > 0$ such that $S(x_0, \delta) \subseteq M$. Therefore,

$$f(S(x_0, \delta)) \subseteq f(M) \subseteq S(f(x_0), \epsilon).$$

That is,

$$d_1(x, x_0) < \delta \text{ implies that } d_2(f(x), f(x_0)) < \epsilon.$$

Theorem 4.1.7 Let (X, d_1), (Y, d_2) and (Z, d_3) be metric spaces, and x_0 be a point in X. If $f : X \to Y$ is a mapping continuous at x_0, and $g : Y \to Z$ is a mapping continuous at $f(x_0)$, then the composition mapping $g \circ f : X \to Z$ is continuous at x_0.

Proof. Let $\epsilon > 0$ be a real number. Let $f(x_0) = y_0$. Then there exists a real number $\delta > 0$ such that

$$d_2(y, y_0) < \delta \quad \text{implies that } d_3(g(y), g(y_0)) < \epsilon \text{ for all } y \text{ in } Y.$$

Again since f is continuous at x_0, there is a real number $\eta > 0$ such that

$$d_1(x, x_0) < \eta \quad \text{implies that } d_2(f(x), f(x_0)) < \delta \text{ for all } x \text{ in } X.$$

Therefore,

$$d_1(x, x_0) < \eta \text{ implies that } d_3(g(f(x)), g(f(x_0))) < \epsilon \text{ for all } x \text{ in } X.$$

Theorem 4.1.8 Let (X, d_1) and (Y, d_2) be metric spaces. Let A be a subset of X, and x_0 be an adherent point of A. If $f : X \to Y$ is a mapping continuous at x_0, then $f(x_0)$ is an adherent point of $f(A)$, that is, $f(x_0) \in \overline{f(A)}$.

Proof. Let N be a neighbourhood of $f(x_0)$ in Y. Since f is continuous at x_0, there exists a neighbourhood M of x_0 such that $f(M) \subseteq N$. Since x_0 is an adherent point of A, $M \cap A \neq \emptyset$. This means that $f(M) \cap f(A)$ is non-empty. Thus, $N \cap f(A)$ is non-empty, and $f(x_0)$ is an adherent point of $f(A)$.

Examples 4.1.9

1. Let $X = \mathbb{R} - (]-1, 0[\cup]0, 1[)$, and $d(x, y) = |x - y|$ for all $x, y \in X$ be the usual metric on X. Let (X, d') be the discrete metric space. Let $i : (X, d) \to (X, d')$ be the identity mapping: $i(x) = x$. Then i is continuous at 0 only. At rest of the points i is not continuous. Extend this example in such a way that i is not continuous at a countable number of points.

2. Let (X, d) be a metric space, then the identity mapping $i : X \to X$ is continuous at every point of X.

3. Let (X, d) be a discrete metric space, and (Y, d') be any metric space. Then every mapping $f : X \to Y$ is continuous at every point of X.

EXERCISE

1. Let $f : l_2 \to \mathbb{R}$ be a mapping defined by $f(\langle x_n \rangle) = x_1$ for all $\langle x_n \rangle$ in l_2. Show that $\lim_{x \to a} f(x) = a_1$, where $a = \langle a_n \rangle \in l_2$.

2. Show that the mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by
$$f(x) = \begin{cases} x & \text{if } \|x\| \leq 1, \\ \dfrac{x}{\|x\|^2} & \text{if } \|x\| > 1 \end{cases}$$
is continuous at every point.

4.2 Continuous Mappings

Definition 4.2.1 Let (X, d_1) and (Y, d_2) be metric spaces. Then a mapping $f : X \to Y$ is said to be continuous on X if it is continuous at every point of X.

Examples 4.2.2

1. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be any constant mapping, that is, $f(x) = y \in Y$ for all $x \in X$. Then f is continuous on X.

2. The identity mapping from a metric space to itself is continuous.

3. Let (X, d) be a metric space. Equip $X \times X$ with metric ρ defined by
$$\rho((x, y), (x', y')) = d(x, x') + d(y, y').$$
Then the mapping $f : X \times X \to \mathbb{R}$ defined by $f(x, y) = d(x, y)$ is continuous on $X \times X$. This is proved using Lemma 3.1.6.
$$|f(x, y) - f(x', y')| = |d(x, y) - d(x', y')| \leq d(x, x') + d(y, y')$$
$$= \rho((x, y), (x', y')).$$

4. The mappings $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x + y$, and $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x, y) = xy$ are continuous. Let $(x, y) \in \mathbb{R}^2$ be arbitrary point of $\mathbb{R}^2$. Let $\langle (x_n, y_n) \rangle$ be a sequence in $\mathbb{R}^2$ converging to (x, y). Then, by Theorem 3.3.2, the sequence $\langle x_n \rangle$ converges to x in $\mathbb{R}$, and $\langle y_n \rangle$ converges to y in $\mathbb{R}$. Then the sequence $\langle (x_n + y_n) \rangle$ converges to $x + y$, that is, the sequence $\langle f(x_n, y_n) \rangle$ converges to $f(x, y)$. Therefore, f is continuous. Similarly, g is continuous on $\mathbb{R}^2$ (See Theorem 4.2.9).

   There are several conditions to equivalent the notion of a continuous mapping.

Theorem 4.2.3 Let (X, d_1) and (Y, d_2) be metric spaces. A mapping $f : X \to Y$ is continuous on X if and only if the inverse image of every open set in Y is an open set in X.

Proof. Suppose that f is continuous on X, and G is an open set in Y. We need to show that $f^{-1}(G)$ is an open set in X. Let x_0 be any point in $f^{-1}(G)$. Then $f(x_0) \in G$. Since

G is open in Y, there exists a real number $\epsilon > 0$ such that $S(f(x_0), \epsilon) \subseteq G$. Since f is continuous at x_0, there exists a real number $\delta > 0$ such that

$$f(S(x_0, \delta)) \subseteq S(f(x_0), \epsilon).$$

This means that$f(S(x_0, \delta)) \subseteq G$, so that

$$S(x_0, \delta) \subseteq f^{-1}(f(S(x_0, \delta)) \subseteq f^{-1}(G).$$

Since x_0 is an arbitrary point, $f^{-1}(G)$ is an open set. Conversely, assume that the inverse image under f of every open set in Y is an open set in X. Let x_0 be a point in X. Let $\epsilon > 0$ be a real number. The open sphere $S(f(x_0), \epsilon)$ in Y is an open set in Y. By our assumption, $f^{-1}(S(f(x_0), \epsilon))$ is an open set in X. Now, $x_0 \epsilon f^{-1}(S(f(x_0), \epsilon))$. Thus, there exists a real number $\delta > 0$ such that $S(x_0, \delta) \subseteq f^{-1}(S(f(x_0), \epsilon))$. This means that $f(S(x_0, \delta)) \subseteq S(f(x_0), \epsilon)$. Therefore, f is continuous at x_0. Since x_0 is arbitrary, the proof is complete.

Theorem 4.2.4 Let (X, d_1) and (Y, d_2) be metric spaces. Then a mapping $f : X \to Y$ is continuous on X if and only if the inverse image of every closed set in Y is a closed set in X.

Proof. Let f be continuous on X, and F be a closed set in Y. Then $Y - F$ is open in Y. By Theorem 4.2.3, $f^{-1}(Y - F)$ is open in X. Since $f^{-1}(Y - F) = X - f^{-1}(F)$, $f^{-1}(F)$ is a closed set in X. Conversely, we show that if inverse n age of every closed set is closed under f then f is continuous on X. Let G be an open set in Y. Then $Y - G$ is a closed set in Y. By our assumption, $f^{-1}(Y - G)$ is a closed set in X. Since $f^{-1}(Y - G) = X - f^{-1}(G)$, $f^{-1}(G)$ is an open set in Y. So f is continuous on X by Theorem 4.2.3.

Corollary 4.2.5 Let (X, d) be a metric space, and A be a non-empty subset of X. If $i : X \to X$ be the identity mapping, then its restriction to $A : i_A : A \to X$ is continuous on A.

Proof. Let G be any open set in X. Then

$$i_A^{-1}(G) = i^{-1}(G) \cap A = G \cap A.$$

The following theorem is a direct consequence of Theorem 4.1.7; here is an alternative proof.

Theorem 4.2.6 Let (X, d_1), (Y, d_2) and (Z, d_3) be metric spaces. If $f : X \to Y$ and $g : Y \to Z$ be continuous mappings, then $g \circ f : X \to Z$ is continuous on X.

Proof. Let U be an open set in Z. Then $g^{-1}(U)$ is open in Y, and $f^{-1}(g^{-1}(U))$ is open in X. Since $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$, the assertion follows from Theorem 4.2.3.

Corollary 4.2.7 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. Then f is continuous on X if and only if the inverse image of every open sphere in Y is an open set in X.

Proof. The necessity part is obvious. Conversely, let G be an open set in Y. G is a union of open spheres in Y, and $f^{-1}(G)$ is the union of inverse images of these open spheres.

Corollary 4.2.8 Let d_1 and d_2 be equivalent metrics on a set X. Let (Y, ρ) be a metric space. Let $f : X \to Y$ be a mapping. Then f is continuous on X with respect to the metric d_1 if and only if f is continuous on X with respect to the metric d_2 on X.

Proof. The result follows immediately from theorem 4.2.3 since (X, d_1) and (X, d_2) both have the same open sets.

The next theorem also gives an equivalent condition for the notion of continuity. The proof follows from Theorem 4.1.3 and Theorem 4.1.5.

Theorem 4.2.9 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. Then f is continuous at a point $x_0 \in X$ if and only if for every sequence $\langle x_n \rangle$ in X converging to x_0, the sequence $\langle f(x_n) \rangle$ converges in Y to $f(x_0)$.

As a consequence of the above theorem, it follows that if f is continuous on X, then the image under f of any convergent sequence in X is a convergent sequence in Y. For this reason, sometimes, we say that a continuous mapping preserves the convergent sequences.

Theorem 4.2.10 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. The following conditions are equivalent.

1. f is continuous on X.
2. For every subset A of X, $f(\bar{A}) \subseteq \overline{f(A)}$.
3. For every subset B of Y, $f^{-1}(B^0) \subseteq (f^{-1}(B))^0$.
4. For every subset B of Y, $\overline{f^{-1}(B)} \subseteq f^{-1}(\bar{B})$.

Proof. (1) implies (2). Suppose that f is continuous. Since $\overline{f(A)}$ is closed in Y, $f^{-1}\overline{(f(A))}$ is a closed set in X by Theorem 4.2.4. Since $f(A) \subseteq \overline{f(A)}$,

$$A \subseteq f^{-1}(f(A)) \subseteq f^{-1}(\overline{f(A)}).$$

Therefore, $\bar{A} \subseteq f^{-1}(\overline{f(A)})$ so that $f(\bar{A}) \subseteq \overline{f(A)}$.

(2) implies (3) Let x be a point in X which is not in $(f^{-1}(B))^0$. Then x is not an interior point of $f^{-1}(B)$. Then $x \in \overline{(X - f^{-1}(B))}$. Therefore, $f(x)$ is in $f(\overline{X - f^{-1}(B)})$. By our assumption, it follows that $f(x)$ is in $(\overline{f(X - f^{-1}(B))})$ which is contained in $\overline{Y - B}$. This means that $f(x) \notin B^0$, that is, $x \notin f^{-1}(B^0)$.

(3) implies (4). Let x be a point in X which is not in $f^{-1}(\bar{B})$. Then $f(x) \notin \bar{B}$. Therefore, $f(x) \in Y - \bar{B} = (Y - \bar{B})^0$. Then $x \in f^{-1}((Y - \bar{B})^0) \subseteq [f^{-1}(Y - \bar{B})]^0 = [X - f^{-1}(\bar{B})]^0 \subseteq [X - f^{-1}(B)]^0$. Therefore, $x \notin \overline{f^{-1}(B)}$.

(4) implies (1). Let F be a closed set in Y. Then $\bar{F} = F$, and, by our assumption, $\overline{f^{-1}(F)} \subseteq f^{-1}(\bar{F})$. Therefore, $f^{-1}(F) \subseteq \overline{f^{-1}(F)} \subseteq f^{-1}(F)$, and the assertion follows by Theorem 4.2.4.

Example 4.2.11 Let (X, d) and (Y_i, d_i), $i = 1, 2, \ldots, k$ be metric spaces, and $f_i : X \to Y_i$, $i = 1, 2, \ldots k$ be mappings. Let$Y = \prod_{i=1}^{k} Y_i$ be equipped with product metric ρ. Then the mapping $\phi : X \to Y$ defined by setting $\phi(x) = (f_1(x), f_2(x), \ldots, f_k(x))$ for all x in X is continuous at a point x_0 in X if and only if for each $i = 1, 2. \ldots, k$, f_i is continuous at x_0. Indeed, if $\langle x_n \rangle$ is any sequence in X converging to x in X, then, by Theorem 4.2.9, the sequence $\langle \phi(x_n) \rangle$ converges to $\phi(x_0)$ in Y if and only if the sequence $\langle f_i(x_n) \rangle_{n=1}^{\infty}$ converges $f_i(x_0)$ for each $i = 1, 2, \ldots, k$ by Theorem 3.3.1.

Theorem 4.2.12 Let (X, d_1) and (Y, d_2) be metric spaces, and A be a subspace of X. Let x_0 be a point in A. If a mapping $f : X \to Y$ is continuous at x_0, then the restriction of f to A, viz., $f_A : A \to Y$ is continuous at x_0.

Proof. Since $f_A = f \circ i_A$, where i_A is the inclusion mapping from A to X which is continuous by Corollary 4.2.5, f_A is continuous at x_0 by Theorem 4.1.7.

It should be noted that the restriction of a mapping f defined on a metric space X to a subspace of X may be continuous without f being continuous at any point of X.

Example 4.2.13 Let $\chi_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ be the characteristic function. Then $\chi_{\mathbb{Q}}$ is not continuous at any point of R. Since the restriction $\chi_{\mathbb{Q}}|_{\mathbb{Q}}$ is a constant function having the value 1 at every point of $\mathbb{Q}$, $\chi_{\mathbb{Q}}|_{\mathbb{Q}}$ is a continuous function. Also, $\chi_{\mathbb{Q}}|_{\mathbb{R}-\mathbb{Q}}$ is continuous on $\mathbb{R} - \mathbb{Q}$.

Now the following definition makes sense.

Definition 4.2.14 Let (X, d_1) and (Y, d_2) be metric spaces, and A be a subspace of X. Let $f : X \to Y$ be a mapping. Then f is said to be continnuous at a point $x_0 \in A$ relative to A if the restriction $f|A$ is continuous at x_0.

In the case when f is continuous at a point $x_0 \in A$ relative to A, x_0 is called a point of continuity of f relative to A otherwise x_0 is called a point of discontinuity of f relative to A. f is said to be continuous on A relative to A if the restriction $f_A : A \to Y$ is continuous on A.

The following theorem is merely a restatement of Theorem 4.1.6 in the context of a relative metric, and the proof is left to the reader.

Theorem 4.2.15 Let (X, d_1) and Y, d_2 be metric spaces, and A be a subset of X. Let $f : X \to Y$ be a mapping. Then the following statements are equivalent.

1. f is continuous at $x_0 \in A$ relative to A.
2. For each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that
$$f(S(x_0, \delta, d_1) \cap A) \subseteq S(f(x_0), \epsilon, d_2).$$
3. The sequence $\langle f(x_n) \rangle$ converges to $f(x_0)$ in Y if the sequence $\langle x_n \rangle$ in A converges in A to x_0.

Example 4.2.16 Let (X, d_1) and (X, d_2) be metric spaces. Then the identity mapping $i : (X, d_1) \to (X, d_2)$ is continuous if and only if d_1 is finer than d_2. Also note that if $f : (X, d_1) \to (X, d_2)$ is a continuous mapping, then making finer the metric d_1 preserves the continuity.

Theorem 4.2.17 Let (X, d) be a metric space, and A be a non-empty subset of X. Then the mapping $f : X \to \mathbb{R}$ given by $f(x) = d(x, A)$ for all $x \in X$ is continuous.

Proof. Let x and x' be any two elements in X. Then for each $a \in A$, by the triangle inequality, $d(x, a) \leq d(x, x') + d(x', a)$, so that

$$\begin{aligned} d(x, A) &= \inf\{d(x, a) : a \in A\} \leq d(x, x') + \inf\{d(x', a) : a \in A\} \\ &= d(x, x') + d(x', A). \end{aligned}$$

This means that $d(x, A) - d(x', A) \leq d(x, x')$, and interchanging x and x', we obtain

$$|d(x, A) - d(x', A)| \leq d(x, x').$$

This proves the continuity of f.

Corollary 4.2.5 is a particular case of the more general result which is as follows.

Theorem 4.2.18 Let $f : (X, d_1) \to (Y, d_2)$ be a continuous mapping, and A be a subspace of X. Then the restriction $f_A : A \to Y$ is continuous on A.

Proof. The proof follows from Theorem 4.2.12.

Example 4.2.19 Let C denote the unit circle in the complex plane with subspace metric, and consider $[0, 1[$ as a subspace of the real line. Define $f : [0, 1[\to C$ by $f(x) = e^{2\pi ix}$. Now let S be any open segment (arc without end points) in C which does not contain the complex number 1, then $f^{-1}(S)$ is an open interval of the form $]a, b[, 0 < a < b < 1$, if S contains 1, then $f^{-1}(S)$ has the form $[0, a[\bigcup]b, 1[, 0 < b < 1$, which is also an open set in $[0, 1[$. Thus, f is continuous. However, the inverse mapping f^{-1} of f is not continuous. The inverse image under f^{-1} of an open set $[0, 1/2[$ in $[0, 1[$ is the set$\{z \in C : 0 \leq \arg(z) < \pi\}$ which is not open in C (Fig. 4.1).

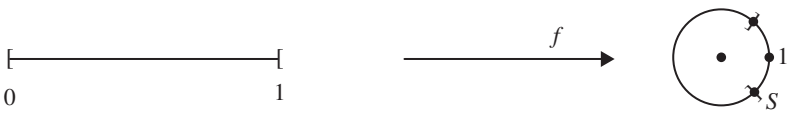


Fig. 4.1

The following theorem discusses continuity of projection mappings.

Theorem 4.2.20 Let (X_1, d_1) and (X_2, d_2) be metric spaces, and $X_1 \times X_2$ be their product metric space. Then the projection mappings $\pi_i : X_1 \times X_2 \to X_i$ defined by $\pi_i(x_1, x_2) = x_i$, $i = 1, 2$ are continuous.

Proof. The continuity of π_1 follows from the following:

$$d_1(\pi_1(x, y), \pi_1(x_0, y_0)) = d_1(x, x_0) \leq \max\{d_1(x, x_0), d_2(y, y_0)\} = d\left((x, y), (x_0, y_0)\right).$$

Similarly, π_2 is continuous.

Theorem 4.2.21 Let V be a normed vector space with norm ϕ, and d be the metric induced by ϕ. Then $\phi : V \to [0, \infty[$, where $[0, \infty[$ is given the usual metric, is continuous.

Proof. Let x and y be any two vectors in V. Then $\phi(x) = \phi(x - y + y) \leq \phi(x - y) + \phi(y)$. Thus, $\phi(x) - \phi(y) \leq \phi(x - y)$. Interchanging x and y, it follows that

$$|\phi(x) - \phi(y)| \leq \phi(x - y) = d(x, y).$$

Therefore, ϕ is continuous.

Theorem 4.2.22 A linear transformation $T : \mathbb{R}^m \to \mathbb{R}^k$ is continuous.

Proof. Let $\langle x_n \rangle$ be sequence in $\mathbb{R}^m$ that converges to x in $\mathbb{R}^m$. Here

$$x_n = (x_{n1}, x_{n2}, \ldots, x_{nm}) = x_{n1}e_1 + x_{n2}e_2 + \cdots + x_{nm}e_m,$$

$e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $\ldots$, $e_m = (0, 0, \ldots, 1)$ are in $\mathbb{R}^m$, $x = (a_1, \ldots, a_m) = a_1e_1 + \cdots + a_me_m$ as usual. We show that the sequence $\langle T(x_n) \rangle$ converges to $T(x)$ in $\mathbb{R}^k$. Since T is a linear transformation

$$\begin{aligned} T(x_n) &= T(x_{n1}e_1 + \cdots + x_{nm}e_m) \\ &= x_{n1}T(e_1) + \cdots + x_{nm}T(e_m). \end{aligned}$$

Hence, by Theorem 3.3.1, the limit of the sequence $\langle T(x_n)\rangle$ is $(a_1T(e_1), a_2T(e_2), \ldots, a_mT(e_m)) = T(a_1e_1 + \cdots + a_me_m) = T(x)$.

Theorem 4.2.23 Let (X, d) be a metric space, and $f : X \to \mathbb{R}$ be a continuous mapping. Let c be a non-negative real number. Then the sets $\{x \in X : f(x) \leq c\}$ and $\{x \in X : f(x) \geq c\}$ are closed in X.

Proof. Let $A = \{x \in X : f(x) \leq c\}$. Let x_0 be a limit point of A. Then there exists a sequence $\langle x_n\rangle$ in A which converges in X to x_0. Since f is continuous, the sequence $\langle f(x_n)\rangle$ converges in $\mathbb{R}$ to $f(x_0)$. Since $f(x_n) \leq c$ for all $n \in \mathbb{Z}_+$, $f(x_0) \leq c$. Thus, $x_0 \in A$, and A is closed. Similarly $\{x \in X$ if $(x) \geq c\}$ is closed.

With the assumption of the above theorem, it follows that the sets $\{x \in X : f(x) > c\}$ and $\{x \in X : f(x) < c\}$ are open sets in X. In particular, the set $\{x \in X : f(x) = 0\}$ is a closed set.

EXERCISE

1. Let $f : (X, d_1) \to (Y, d_2)$ be a continuous mapping, and $A \subseteq X$. Show that if a point x in X is a limit point of A, it is necessary that $f(x)$ is a limit point of $f(A)$.

2. Let (X, d_1) and (Y, d_2) be two metric spaces, and f and g be two continuous mappings from (X, d_1) into (Y, d_2). Prove that if $f(x) = g(x)$ for all x in $A \subseteq X$, then $f(x) = g(x)$ for all x in $\bar{A}$. Thus, if $B = \{x : f(x) = g(x)\}$, then B is a closed set.

3. Give examples of metric spaces (X, d_1) and (Y, d_2), and a mapping $f : X \to Y$ which is not continuous. For this mapping f, find

   (a) an open set G in Y such that $f^{-1}(G)$ is not an open set in X.

   (b) a closed set F in Y such that $f^{-1}(F)$ is not closed in X.

   (c) a set A in X such that $f(\bar{A}) \not\subseteq \overline{f(A)}$.

4. Let $\mathcal{C}[a, b]$ be the metric space in Example 2.1.10. Let $F : \mathcal{C}[a, b] \to \mathbb{R}$ be the mapping defined by $F(f) = \int_a^b f(t)dt$. Then F is continuous.

5. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a continuous mapping, then the graph of f defined by

   $$G = \{(x, y) \in X \times Y : y = f(x)\}$$

   is a closed set in the product metric space $X \times Y$.

6. Let $f : (X, d_1) \to (Y, d_2)$ be a continuous mapping, and A be a subset of X. If $x \in A$, and $f(x)$ is an interior point of $f(A)$, then prove that x is an interior point $f^{-1}(f(A))$. Use the following example to show that x is not necessarily an interior point of A. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = |x|$, and take A to be the set of rational numbers in $[0, 1]$ and irrational numbers in $]-1, 0[$.

7. Let (X, d) and (Y, ρ) be metric spaces, and A and B be subsets of X such that $A \cup B = X$. Let $f : (A, d_A) \to (Y, \rho)$, $g : (B, d_B) \to (Y, \rho)$ be continuous mappings, and $f(x) = g(x)$ for all x in $A \bigcap B$. Then the mapping $h : X \to Y$ defined by

$$h(x) = \begin{cases} f(x) & \text{if } x \in A, \\ g(x) & \text{if } x \in B \end{cases}$$

is not necessarily continuous. If, however, A and B are both open, or both closed, then prove that h is continuous.

8. Show that a mapping f from a metric space X into a metric space Y is continuous at every isolated point of X. Describe the continuous mappings from Y into X when X is discrete.

9. Let $f : \mathbb{N} \to \mathbb{Q}$ be a bijective mapping. Show that f is continuous, but f^{-1} is not continuous at any point of $\mathbb{Q}$.

10. Let X and Y be metric spaces, and A be a subset of X. If a mapping $f : A \to Y$ is continuous at a point $x_0 \in A$ relative to A, then f is continuous at x_0 relative to any subset B such that $x_0 \in B \subseteq A$.

11. Let X and Y be metric spaces, and E and F be non-empty disjoint closed sets in X. Show that $d(x, E) + d(x, F) > 0$ for all x in X. Therefore, we can define a mapping f by

$$f(x) = d(x, E)/[d(x, E) + d(x.F)].$$

Show that the mapping $f : X \to \mathbb{R}$ is continuous, and $E = \{x \in X : f(x) = 0\}$, and $F = \{x \in X : f(x) = 1\}$. Also show that there exists a real number M such that

$$|f(x) - f(y)| \leq Md(x, y)$$

if and only if $d(E, F) > 0$. If we denote f by $f_{E,F}$, then show that

$$E_1 \subseteq E_2 \text{ implies that } f_{E_1,F}(x) \geq f_{E_2,F}(x) \text{ for all } x \in X,$$

that is, $f_{E,F}$ is monotone decreasing with respect to first variable E. Similarly, show that $F_1 \subseteq F_2$ implies that $f_{E,F_1}(x) \leq f_{E,F_2}(x)$ for all $x \in X$, that is, $f_{E,F}$ is monotone increasing with respect the second variable F.

12. Show that the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R} - \mathbb{Q} \end{cases}$$

is continuous at the point $x = 0$ and continuous at no other point. Likewise, show that the function $g : \mathbb{R} \to \mathbb{R}$ defined by

$$g(x) = \begin{cases} x - x^2 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{R} - \mathbb{Q} \end{cases}$$

is continuous at $x = 0$ and 1 only and at no other point. Also construct a real valued function on $\mathbb{R}$ that is continuous at every point of the cantor set C, and discontinuous at every point of $\mathbb{R} - C$.

13. Let X and Y be metric spaces, and $B \subseteq A \subseteq X$. Let $a_0 \in \overline{A} \cap \overline{B}$. Let $f : A \to Y$ be a mapping. Show that if $\lim_{\substack{x \to a_0 \\ x \in A}} f(x) = y_0 \in Y$, then $\lim_{\substack{x \to a_0 \\ x \in B}} f(x) = y_0$. Verify that if B contains a set of the form $S(a_0, \epsilon) \cap A$, then if $\lim_{\substack{x \to a_0 \\ x \in B}} f(x) = y_0$, then $\lim_{\substack{x \to a_0 \\ x \in A}} f(x) = y_0$. Also show that if $a_0 \in \overline{(A - B)}$, then $\lim_{\substack{x \to a_0 \\ x \in A}} f(x) = y_0$ if and only if $\lim_{\substack{x \to a_0 \\ x \in B}} f(x) = \lim_{\substack{x \to x_0 \\ x \in A - B}} f(x) = y_0$.

14. Let X be a separable metric space, and $f : X \to \mathbb{R}$ be a mapping. Show that the set of points $a \in X$ such that $\lim_{\substack{x \to a \\ x \neq a}} f(x)$ exists and is distinct from $f(a)$ is at most denumerable. For every pair of rational numbers p,q such that $p < q$, the set of points $a \in X$ such that

$$f(a) \leq p < q \leq \lim_{\substack{x \to a \\ x \neq a}} f(x)$$

is at most denumerable. Similarly, the set of points $a \in X$ such that

$$\lim_{\substack{x \to a \\ x \neq a}} f(x) \leq p < q \leq f(a)$$

is at most denumerable.

15. Recall that $\mathbb{Q}_+$ is the set positive rational numbers. Let a be a positive irrational number, and $f_a : \mathbb{Q}_+ \to]0, a[$ be a function such that $x - f_a(x))$ is an integral multiple of a. Show that f_a is injective and continuous, and that $f_a(\mathbb{Q}_+)$ is dense in $]0, a[$. Deduce from that result that there exists a bijective continuous mapping of Q onto itself.

16. Let (X, d) be a metric space, and $f : X \to X$ be a continuous mapping. Prove that the set $\{x \in X : f(x) = x\}$ is a closed set in X.

17. Let (X, d) be a metric space. A subset A of X is called a G_δ-set in X if there exists a sequence $\langle G_n \rangle$ of open sets in X such that $A = \bigcap_{n=1}^{\infty} G_n$. A subset B of X is called an F_σ-set if there exists a sequence $\langle F_n \rangle$ of closed sets in X such that $B = \bigcup_{n=1}^{\infty} F_n$. Now prove the following :

    (a) The complement in X of a G_δ-set is an F_σ-set, and conversely the complement of an F_σ-set in X is a G_δ-set in X. (In this sense the notions of a G_δ-set and F_σ-set are known as dual of each other.)

    (b) Every closed subset of X is a G_δ-set, and every open set in X is an F_σ-set (Hint: If F is a non-empty closed set, $f(x) = d(x, F)$ for all x in X. Then $F = \{x \in X : f(x) = 0)\}$, while the set of the form $\{x \in X : f(x) < c\}$, where c is a non-negative real number, is open by Theorem 4.2.23.)

18. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational,} \\ \dfrac{1}{n} & \text{if } x \text{ is a rational number } \frac{m}{n} \text{ } (m \text{ and } n \text{ are integers} \\ & \text{having no common factors, and } n > 0). \end{cases}$$

    Prove that f is continuous at every irrational point and discontinuous at every rational point.

19. Prove the impossibility of constructing a real valued function on $[0,1]$ which is continuous at every rational point, and is discontinuous at every irrational point.

20. Find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x)$ is irrational if x is rational, and $f(x)$ is a rational if x is irrational.

21. Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ be a mapping defined by $f(n,m) = n/m$. If $\mathbb{N} \times \mathbb{N}$ is equipped with the metric
$$\rho((n_1,n_2),(m_1,m_2)) = \left|\frac{n_1}{n_2} - \frac{m_1}{m_2}\right|$$
and $\mathbb{Q}$ with the usual metric, then show that f is continuous.

22. Let (X,d) be a metric space, $\emptyset \neq A \subseteq X$, and $f : X \to \mathbb{R}$ be a mapping defined by $f(x) = d(x,A)$ for all x in X. Show that $f(x) = 0$ if and only if $x \in \bar{A}$.

23. Show that the mapping $f : \mathbb{R}^2 \to \mathbb{R}$ defined by
$$f(x,y)\begin{cases}\dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0),\\ 0 & \text{if } (x,y) = (0,0)\end{cases}$$
is not continuous at (0,0).

24. Let A and B be disjoint closed sets in a metric space X. Show that there exists a continuous real valued mapping f such that $f(A) = \{1\}$, $f(B) = \{-1\}$, and $-1 < f(x) < 1$ for all x in $X - (A \cup B)$.

25. Recall the usual definition of derivative of a real valued function f of a real variable. If the limit $\lim_{\Delta x \to 0}(f(x+\Delta x) - f(x))/\Delta x$ exists, then f is said to be differentiable at x, and the value of the limit denoted by $f'(x)$, is called the derivative of f at x. Now recall Rolle's theorem and Lagrange theorem:

    Theorem (Rolle): Let f be continuous real valued function defined on the closed interval $[a,b]$, $a < b$, differentiable at each point of $]a,b[$, and $f(a) = f(b)$. Then there exists a point c in $]a,b[$ such that $f'(c) = 0$.

    Theorem (Lagrange): Let f be a continuous real valued function defined on the closed interval $[a,b]$, $a < b$, and differentiable at each point of $]a,b[$. Then there exists a point $c \in]a,b[$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.

    Prove that the Rolle's Theorem and Lagrange's Theorem are equivalent by using the function:
$$\begin{vmatrix} f(x) & f(a) & f(b) \\ x & a & b \\ 1 & 1 & 1 \end{vmatrix}$$

26. Let (X,d_1) be a metric space, and $\langle A_n\rangle$ be a sequence of subsets of X such that $X = \bigcup_{n=1}^{\infty} A_n$, and $A_n \subseteq A_{n+1}^0$ for each n. Let (Y,d_2) be a metric space, and $f : X \to Y$ be such that the restrictions $f|A_n : A_n \to Y$ are continuous with respect to the induced metrics on A_n's. Show that f itself is continuous.

4.3 Isometry Mappings

Recall the notion of isomorphism between partially ordered sets. The corresponding notion in metric spaces is known as isometry.

Definition 4.3.1 Let (X, d_1) and (Y, d_2) be two metric spaces. A mapping $f : X \to Y$ is called isometric if it preserves the distance, that is, $d_2(f(x), f(y)) = d_1(x, y)$ for all x, y in X. An isometric mapping f of X onto Y is called an isometry.

It is easy to see that an isometry is injective, and that inverse and compositions of isometries are isometries. Also identity mapping from a metric space to itself is an isometry.

Examples 4.3.2

1. Let a be a fixed point in $\mathbb{R}^n$. Then the translation $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by $f(x) = x+a$ is an isometry. Infact, if we take the metric d_p on $\mathbb{F}^n$ in Example 2.1.8, then the translation $x \to x + a$ is an isometry. Also the mapping $f : \mathbb{C}^n \to \mathbb{R}^n$ defined by $f(z_1, \ldots, z_n) = (\operatorname{Re} z_1), \operatorname{Im} z_1, \ldots, \operatorname{Re} z_n, \operatorname{Im} z_n)$ is an isometry. The reflection mapping $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (-x, y)$ is an isometry. The rotation mapping $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x, y) = (-x_1 \sin\theta + x_2 \cos\theta, x_1 \cos\theta + x_2 \sin\theta)$$

   is also an isometry.

2. Let $\mathbb{R}$ be the real line, and X be the set of all functions of the form $f(x) = \alpha x$, where $\alpha \in \mathbb{R}$. The mapping $d^* : X \times X \to \mathbb{R}$ defined by $d^*(f, g) = |\alpha - \beta|$, where $f(x) = \alpha x$, and $g(x) = \beta x$ for $\alpha, \beta \in \mathbb{R}$, is a metric on X. Then the mapping $T : X \to \mathbb{R}$ defined by $T(f) = \alpha$, where $f(x) = \alpha x$, $\alpha \in \mathbb{R}$, is an isometry:

$$d(T(f), T(g)) = d(\alpha, \beta) = |\alpha - \beta| = d^*(f, g),$$

   where d is the usual metric on $\mathbb{R}$.

3. Let X and Y be two metric spaces having the same cardinal number, and both equipped with discrete metric. Then any one-to-one correspondence between X and Y is an isometry.

   There are several properties which are preserved under isometries. The next two theorems are immediate consequences of the definition of an isometry.

Theorem 4.3.3 Let (X, d_1) be a bounded metric space, and (Y, d_2) be another metric space. Let $f : X \to Y$ be an isometry. Then (Y, d_2) is bounded.

Theorem 4.3.4 Let (X, d_1) be a discrete metric space, and (Y, d_2)be a metric space. If $f : X \to Y$ is an isometry, then (Y, d_2) is a discrete metric space.

Theorem 4.3.5 Separability is preserved under isometries.

Proof. Let (X, d_1) be a separable metric space, and (Y, d_2) be another metric space. Let $f : X \to Y$ be an isometry. We show that (Y, d_2) is separable. Assume that A is a countable and dense set in X. Since f is bijective, $f(A)$ is countable. Let y be a point in $Y - f(A)$. Then there exists a point $x \in X$ such that $f(x) = y$. Let $\epsilon > 0$ be a real number. Then there exists a point x' in A such that x' is in $S(x, \epsilon)$. Since $f(x') \in f(A)$,

and $d_2(f(x'), y) = d_2(f(x'), f(x)) = d_1(x', x) < \epsilon$, it follows that y is a limit point of $f(A)$, and so $f(A)$ is dense in Y.

A metric space (X, d_1) is isometric to a metric space (Y, d_2), written as $(X, d_1) \sim (Y, d_2)$ if there is an isometry from X onto Y. The relation $\sim$ is an equivalence relation in a set of metric spaces.

Theorem 4.3.6 Let $\mathcal{A}$ be a set of metric spaces. Then the relation $\sim$ is an equivalence relation on $\mathcal{A}$.

Proof. Let (X, d) be a metric space in $\mathcal{A}$. Since the identity mapping $i : X \to X$ is an isometry, $(X, d) \sim (X, d)$, and so the relation $\sim$ is reflexive. Let (X, d_1) and (Y, d_2) be metric spaces in $\mathcal{A}$ such that $(X, d_1) \sim (Y, d_2)$. Then there is an isometry from (X, d_1) onto (Y, d_2). Now f^{-1} is an isometry from (Y, d_2) onto (X, d_1). Therefore, $\sim$ is symmetric. Since the composition of isometries is an isometry, the transitivity is obvious. Thus, the relation $\sim$ is an equivalence relation.

In view of the above theorem, the relation $\sim$ on any set $\mathcal{A}$ of metric spaces partitions $\mathcal{A}$ into equivalence classes.

The following theorem may be used to transport the metric of a metric space (X, d) to a non-empty set Y.

Theorem 4.3.7 Let (X, d_1) be a metric space, Y be a non-empty set. Let $f : X \to Y$ be a bijective mapping. Then there is a metric d_2 on Y such that (X, d_1) and (Y, d_2) are isometric.

Proof. The mapping $d_2 : Y \times Y \to \mathbb{R}$ defined by

$$d_2(f(x), f(y)) = d_1(x, y) \text{ for all } x, y \text{ in } X$$

is a metric on Y as can be easily verified. From the definition of d_2, f is an isometry from X onto Y, and so (X, d_1) and (Y, d_2) are isometric.

Theorem 4.3.8 Every isometry is a continuous mapping.

Proof. Obvious.

Theorem 4.3.9 Let (X_1, d_1) and (X_2, d_2) be two metric spaces, and $a_1 \in X_1$ and $a_2 \in X_2$. Then the mapping $x_1 \to (x_1, a_2)$ (respectively, $x_2 \to (a_1, x_2)$) is an isometry of X_1 (respectively, X_2) onto the closed subspace $X_1 \times \{a_2\}$ (respectively, $\{a_1\} \times X_2$) of the product metric space $(X_1 \times X_2, d)$.

Proof. Let $f : X_1 \to X_1 \times \{a_2\}$ be a mapping defined by $f(x) = (x, a_2)$. Then f is onto, and $d(f(x), f(y)) = d((x, a_2), (y, a_2)) = \max\{d_1(x, y), d_2(a_2, a_2)\} = d_1(x, y)$.

Theorem 4.3.10 Metric spaces (X_i, d_i), $i = 1, 2, \ldots, n$ are separable metric spaces if and only if their product metric space (X, d) is separable.

Proof. Assume that (X_i, d_i), $i = 1, 2, \ldots, n$ are separable. Then for each i, there exists a countable and dense set A_i in X_i. Then the set $A = A_1 \times A_2 \times \cdots \times A_n$ is countable, and $\overline{A} = \overline{A_1} \times \overline{A_2} \times \ldots \overline{A_n}$ by theorem 2.15.5. Since $\overline{A_i} = X_i$, $\overline{A} = X_1 \times X_2 \times \cdots \times X_n$. Conversely, let (X, d) be a separable metric space. Let $a_i \in X_i$, $i = 1, 2, \ldots, n$. By Theorem 2.15.6, $\{a_1\} \times \cdots \times \{a_{i-1}\} \times X_i \times \{a_{i+1}\} \times \cdots \times \{a_n\}$ is a closed subset of X, and, by Corollary 2.13.13, this closed subset is a separable subspace of X. By Theorem 4.3.9, the mapping $f_i : X_i \to \{a_1\} \times \cdots \times \{a_{i-1}\} \times X_i \times \{a_{i+1}\} \times \cdots \times \{a_n\}$ defined by$f_i(x) = (a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n)$ is an isometry. Therefore, X_i is separable by Theorem 4.3.5.

Example 4.3.11 $\mathbb{R}^n$ is separable since $\mathbb{R}$ is separable. Also, $\mathbb{C}^n$ is separable.

Theorem 4.3.12 The metric space l_p, $1 \leq p < \infty$, is isometric to a nowhere dense subset of itself.

Proof. Let E be a subset of l_p consisting all points whose first co-ordinate is zero. Let $f : l_p \to E$ be a mapping defined by

$$f(\langle x_1, x_2, \dots \rangle) = \langle 0, x_1, x_2, \dots \rangle.$$

The mapping f is clearly bijective, and it is an isometry since for $x = \langle x_n \rangle$ and $y = \langle y_n \rangle$ in l_p,

$$\begin{aligned} d_p(x, y) &= \left[\sum_{i=1}^{\infty} |x_i - y_i|^p \right]^{\frac{1}{p}} = \left[0 + \sum_{i=2}^{\infty} |x_{i-1} - y_{i-1}|^p \right]^{\frac{1}{p}} \\ &= d_p(f(x), f(y)). \end{aligned}$$

To show that E is nowhere dense subset of l_p, consider an open sphere $S(x, \epsilon)$ centered at $x = \langle x_1, x_2, \dots \rangle$ in l_p. Now we may choose a non-zero real number a such that $|x_1 - a| < \frac{\epsilon}{2}$. The point $y = \langle a, x_2, x_3, \dots \rangle$ is in l_p. Let $\epsilon^* = \min\{|a|, \frac{\epsilon}{2}\}$. We show that $S(y, \epsilon^*) \subseteq S(x, \epsilon)$, and $S(y, \epsilon^*) \cap E = \emptyset$. Let z be any point in $S(y, \epsilon^*)$. Then

$$d_p(x, z) \leq d_p(x, y) + d_p(y, z) < \frac{\epsilon}{2} + \epsilon^* < \epsilon.$$

Thus, $z \in S(x, \epsilon)$. If $S(y, \epsilon^*)$ is not disjoint from E, then a point z in $S(y, \epsilon^*) \cap E$ would have its first co-ordinate zero, and $d_p(y, z)$ would be atleast as large as $|a|$. Thus, $d_p(y, z) \geq |a| \geq \epsilon^* > d_p(y, z)$ which is a contradiction.

EXERCISE

1. Prove that the only isometries of the real line $\mathbb{R}$ are $f(x) = x + a$, and $f(x) = -x + a$, $a \in \mathbb{R}$.
2. Let (X, d) be a metric spaces, Let $\mathcal{H}$ be the set of all closed and bounded subsets of X. Let σ be the Hausdroff metric on $\mathcal{H}$. Then the mapping $f : X \to \mathcal{H}$ defined by $f(x) = \{x\}$ is an isometry.

4.4 Extension Theorems

Theorem 4.4.1 Let f and g be two continuous mappings from a metric space (X, d_1) into a metric space (Y, d_2). Then the set $F = \{x \in X : f(x) = g(x)\}$ is a closed set in X.

Proof. Let y be an element in $X - F$. Then $f(y) \neq g(y)$. Let $d_2(f(y), g(y)) = \epsilon$. Since f and g are continuous at y, there exists a neighbourhood N of y such that

$$d_2(f(x), f(y)) < \frac{\epsilon}{2} \quad \text{and} \quad d_2(g(x), g(y)) < \frac{\epsilon}{2} \text{ for all } x \in N.$$

Then for $x \in N$, $f(x) \neq g(x)$, otherwise we would have $d_2(f(y), g(y)) < \epsilon$ by the triangle inequality. This means that N is contained in $X - F$, and F is a closed set in X.

Theorem 4.4.2 Let f and g be continuous mappings from a metric space (X, d_1) into a metric space (Y, d_2). If $f(x) = g(x)$ for all points x in a dense subset A in X, then $f = g$.

Proof. The set $F = \{x \in X : f(x) = g(x)\}$ is a closed set by the preceding theorem, and this set contains A. Thus, $X = \bar{A} \subseteq \bar{F} = F$, and so $F = X$.

The above theorem is known as the principle of extension of identities.

Theorem 4.4.3 Let f and g be two continuous mappings from a metric space (X, d_1) into $\mathbb{R}$. Then the set $F = \{x \in X : f(x) \leq g(x)\}$ is closed in X.

Proof. Let x_0 be a point in $X - F$. Then $f(x_0) > g(x_0)$. Therefore, there exists a real number α such that $f(x_0) > \alpha > g(x_0)$. Since f is continuous, $f^{-1}(]\alpha, \infty[)$ is an open set in X containing x_0. Similarly, $g^{-1}(]-\infty, \alpha[)$ is an open set in X containing x_0. Hence, $f^{-1}(]\alpha, \infty[) \cap g^{-1}(]-\infty, \alpha[)$ is a neighbourhood of x_0, and for any point y in $f^{-1}(]\alpha, \infty[) \cap g^{-1}(]-\infty, \alpha[)$, $g(y) < \alpha < f(y)$. Therefore, $f^{-1}(]\alpha, \infty[) \cap g^{-1}(]-\infty, \alpha)$ is contained in $X - F$. Since x_0 is arbitrary, $X - F$ is an open set in X.

Theorem 4.4.4 Let f and g be two continuous mappings from a metric space (X, d_1) into $\mathbb{R}$. If $f(x) \leq g(x)$ for all points x in a dense set A in X, then $f(x \leq g(x))$ for all x in X .

Proof. The set $F = \{x \in X : f(x) \leq g(x)\}$ is a closed set by the preceding theorem. Since A is contained in F, $X = \bar{A} \subseteq \bar{F} = F$. Therefore, $F = X$.

The above theorem is known as the principle of extension of inequalities.

Theorem 4.4.5 Let (X, d_1) and (Y, d_2) be metric spaces, and A be a dense subset of X. Let $f : A \to Y$ be a continuous mapping. Then there exists a continuous mapping $\overline{f} : X \to Y$ which coincides with f on A, that is, $\overline{f}(x) = f(x)$ for all x in A if and only if for any $x \in X$, the limit $\lim_{\substack{y \to x \\ y \in A}} f(y)$ exists. The continuous mapping $\overline{f}$ is unique.

Proof. Let $\overline{f}$ be a continuous mapping from X into Y coinciding with f on A. Let $x \in X$. Then $x \in \overline{A}$, and $\lim_{y \to x} \overline{f}(y) = \overline{f}(x)$. Then, by theorem 4.1.5, it follows that $\lim_{\substack{y \to x \\ y \in A}} \overline{f}(y) = \overline{f}(x)$. Since $\overline{f}(y) = f(y)$ for all y in A, $\lim_{\substack{y \to x \\ y \in A}} f(y) = \overline{f}(x)$. This shows that $\lim_{\substack{y \to x \\ y \in A}} f(y)$ exists in Y. Conversely, suppose that for any point $x \in X$ the limit $\lim_{\substack{y \to x \\ y \in A}} f(y)$ exists in Y. Let us define $\overline{f} : X \to Y$ by

$$\overline{f}(x) = \lim_{\substack{y \to x \\ y \in A}} f(y).$$

(We can define $\overline{f}$ in this way since for any point x in $X = \overline{A}$, the limit $\lim_{\substack{y \to x \\ y \in A}} f(y)$ exists by our assumption.) First we show that $\bar{f}$ is continuous on X. Let x_0 be a point in X. Let $\epsilon > 0$ be a real number. Then, by the definition of limit, there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \quad \text{and } x \in A \quad \text{implies that} \quad d_2(f(x), \overline{f}(x_0)) < \frac{\epsilon}{2}.$$

That is,

$$f(S(x_0, \delta) \cap A) \subseteq S(\bar{f}(x_0), \frac{\epsilon}{2}),$$

Now let x' be any point in $S(x_0, \delta)$. Then $x' \in \overline{A \cap S(x_0, \delta)}$, and, by Theorem 4.1.5,

$$\overline{f}(x') = \lim_{\substack{x \to x' \\ x \in A \cap S(x_0, \delta)}} f(x).$$

Therefore,

$$\overline{f}(x') \in \overline{f(A \cap S(x_0, \delta))} \subseteq \overline{S(\overline{f}(x_0), \frac{\epsilon}{2})} \subseteq S(\bar{f}(x_0), \epsilon)$$

which means that $\overline{f}$ is continuous at x_0. Since x_0 is arbitrary, $\overline{f}$ is continuous X on. Now we show that $\overline{f}(x) = f(x)$ for all x in A. Let x_0 be a point in A. Since $\bar{f}$ is continuous at x_0, for each real number $\epsilon > 0$, there exists a real number $\delta_1 > 0$ such that

$$d_1(x, x_0) < \delta_1 \text{ implies that } d_2(\overline{f}(x), \overline{f}(x_0)) < \frac{\epsilon}{2}.$$

Since $\overline{f}(x_0) = \lim_{\substack{x \to x_0 \\ x \in A}} f(x)$, there exists a real number $\delta_2 > 0$ such that

$$d_1(x, x_0) < \delta_2 \quad \text{and } x \in A \quad \text{implies that} \quad d_2(f(x), \overline{f}(x_0)) < \frac{\epsilon}{2}.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then $x \in S(x_0, \delta)$, and $x \in A$ implies that

$$d_2(f(x), \overline{f}(x)) \leq d_2(f(x), \overline{f}(x_0)) + d_2(\overline{f}(x_0), \overline{f}(x)) < \epsilon.$$

Since ϵ is arbitrary, this means that $\overline{f}(x) = f(x)$ for all x in A. Therefore, $\overline{f}$ is a continuous extension of f. Now it remains to show that $\overline{f}$ is unique. Let g be another continuous extension of f. Then $\overline{f}(x) = g(x)$ for all x in A. But then $\overline{f} = g$ by Theorem 4.4.2.

The above theorem is known as the principle of extension of continuous mappings. In general, a continuous mapping on a dense subset A of a metric space X cannot be extended to a continuous mapping on X. For example, let $[0, 1]$ be the subspace of $\mathbb{R}$. Then the open interval $]0, 1[$ is dense in $[0, 1]$. The mapping $f :]0, 1[\to \mathbb{R}$ defined by $f(x) = 1/x$ is continuous, and f cannot be extended to a continuous mapping $\overline{f} : [0, 1] \to \mathbb{R}$ such that $\overline{f}(x) = f(x)$ for all x in $]0, 1[$.

Lemma (Gluing) 4.4.6 Let X and Y be metric spaces.

1. Let $\{G_i\}_{i \in \Lambda}$ be a family of open sets in X that covers X, that is, $\bigcup_{i \in \Lambda} G_i = X$. For each $i \in \Lambda$, let $f_i : G_i \to Y$ be a continuous mapping with the property that $f_i(x) = f_j(x)$ for all $x \in G_i \cap G_j$, and $i, j \in \Lambda$. Then the mapping $f : X \to Y$ defined by $f(x) = f_i(x)$ for all $x \in G_i$ is continuous.

2. Let $\{F_i\}_{i \in \Lambda}$ be a finite family of closed sets in X that covers X, that is, $\bigcup_{i \in \Lambda} F_i = X$. For each $i \in \Lambda$ let $f_i : F_i \to Y$ be a continuous mapping with the property that $f_i(x) = f_j(x)$ for all $x \in F_i \cap F_j$, and $i, j \in \Lambda$. Then the mapping $f : X \to Y$ defined by $f(x) = f_i(x)$ for all $x \in F_i$ is continuous.

Proof. (1) f is clearly well-defined. For any $x \in X$, let Λ', be the set of all indices $i \in \Lambda$ such that $x \in G_i$. Then $f(x)$ is the common value of $f_i(x)$ for all $i \in \Lambda'$. To show that f is continuous, let V be an open set in Y. Since $f_i : G_i \to Y$ is continuous, $f_i^{-1}(V)$ is open in G_i. Since G_i is open in X, $f_i^{-1}(V)$ is open in X by Corollary 2.13.6. We assert that $f_i^{-1}(V) = f^{-1}(V) \cap G_i$. A point $x \in f^{-1}(V) \cap G_i$ if and only if $f(x) \in V$, and $x \in G_i$ if and only if $f_i(x) = f(x) \in V$ if and only if $x \in f_i^{-1}(V)$. That $f^{-1}(V)$ is open in X follows from the equalities:

$$f^{-1}(V) = f^{-1}(V) \cap X = f^{-1}(V) \cap (\bigcup_{i \in \Lambda} G_i) = \bigcup_{i \in \Lambda} (f^{-1}(V) \cap G_i).$$

This completes the proof of the fact that f is continuous.

(2) The proof of (2) is similar.

A continuous real valued bounded mapping defined on a closed subset of a metric space X can be extended to a continuous real valued bounded mapping on X. This fact is known as Tietze-Urysohn theorem.

Theorem 4.4.7 Let A be a closed subset of a metric space (X, d). Let$f : A \to \mathbb{R}$ be a bounded continuous mapping. Then there exists a bounded continuous mapping $\bar{f} : X \to \mathbb{R}$ such that $\bar{f}(x) = f(x)$ for all x in A, and is such that $\inf_{x\in X} \bar{f}(x) = \inf_{x\in A} f(x)$, $\sup_{x\in X} \bar{f}(x) = \sup_{x\in A} f(x)$.

Proof. If f is a constant mapping, then the proof is obvious. Without loss of generality, we can assume that $\inf\{f(y) : y \in A\} = 1$, and $\sup\{f(y) : y \in A\} = 2$; if this is not the case, replace f by a mapping $y \to \alpha f(y) + \beta$, $\alpha \neq 0$. Let $\bar{f} : X \to \mathbb{R}$ be defined by

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in A, \\ \dfrac{\inf_{y\in A}\{f(y)d(x,y)\}}{d(x,A)} & \text{if } x \in X - A. \end{cases}$$

Since $1 \leq f(y) \leq 2$ and $d(x, A) > 0$ for $x \in X - A$, it follows that $1 \leq \bar{f}(x) \leq 2$, and $\bar{f}(x) = f(x)$ for all x in A. Therefore, we prove that $\bar{f}$ is continuous on X. If x_0 is an interior point of A, then the continuity of $\bar{f}$ at x_0 follows from the continuity of f at x_0. Let $\epsilon > 0$ be a real number. Since f is continuous at x_0, there exists a real number $\delta > 0$ such that, $S(x_0, \delta) \subseteq A^0$, and $f(S(x_0, \delta)) \subseteq S(f(x_0), \epsilon)$. Since $\bar{f}$ and f coincides on $S(x_0, \delta)$, it follows that $\bar{f}(S(x_0, \delta)) \subseteq S(\bar{f}(x_0), \epsilon)$ which means that $\bar{f}$ is continuous at x_0. If x is a point in the open set $X - A$, then we can write

$$\bar{f}(x) = \frac{h(x)}{d(x, A)} \quad \text{where } h(x) = \inf_{y\in A}\{f(y)d(x, y)\}.$$

Now $d(x, A)$ is continuous and non-zero (Theorem 2.9.6). Therefore, $\bar{f}$ will be continuous on $X - A$ if h is continuous on $X - A$. Let x be a point in $X - A$. Let $r = d(x, A)$. Then for $d(x, x') \leq \epsilon < r$, we have

$$d(x, y) \leq d(x, x') + d(x', y) \leq d(x', y) + \epsilon.$$

Multiplying $f(y)$ in the above inequality, we have,

$$f(y)d(x, y) \leq f(y)d(x', y) + 2\epsilon$$

since $f(y) \leq 2$. Therefore, $h(x) \leq h(x') + 2\epsilon$. Similarly, $h(x') \leq h(x) + 2\epsilon$ so that

$$|h(x) - h(x')| \leq 2\epsilon$$

which proves the continuity of h. Finally, let x be a boundary point of A. Then $x \in A$. Let $\epsilon > 0$ be a real number. Since f is continuous at x, there exists a real number $\delta > 0$ such that

$$y \in S(x, \delta) \cap A \quad \text{implies that } |f(y) - f(x)| < \epsilon.$$

Let $C = S(x, \delta) \cap A$, and $D = A - C$. If x' is any point in $X - A$ such that $d(x, x') \leq \frac{\delta}{4}$, then for any point $y \in D$

$$d(x', y) \geq d(x, y) - d(x, x') \geq \frac{3\delta}{4}.$$

Therefore,

$$\inf_{y\in D}\{f(y)d(x',y)\} \geq \frac{3\delta}{4}$$

since $f(y) \geq 1$. On the other hand,

$$f(x)d(x',x) \leq 2d(x',x) \leq \frac{\delta}{2}.$$

Since $x \in C$,

$$\inf_{y\in C}\{f(y)d(x',y)\} \leq \frac{\delta}{2}.$$

Therefore,

$$\inf_{y\in A}\{f(y)d(x',y)\} = \inf_{y\in C}\{f(y)d(x',y)\}.$$

But for any $y \in C$,

$$f(x) - \epsilon < f(y) < f(x) + \epsilon$$

and $d(x', A) = \inf_{y\in A}\{d(x', y)\} = \inf_{y\in C}\{d(x', y)\}$ since $x \in C$, $d(x', x) \leq \frac{\delta}{4}$, and for all $y \in D$, $d(x', y) \geq \frac{3\delta}{4}$. Therefore,

$$(f(x) - \epsilon)d(x', A) \leq \inf_{y\in A}\{f(y)d(x', y)\} \leq (f(x) + \epsilon)d(x', A)$$

that is,

$$f(x) - \epsilon \leq \bar{f}(x') \leq f(x) + \epsilon$$

or $|\bar{f}(x') - f(x)| \leq \epsilon$. Thus, $x' \in X - A$, and $d(x, x') \leq \frac{\delta}{4}$ implies that $|\bar{f}(x') - \bar{f}(x)| \leq \epsilon$. If x' is a point in A, and $d(x, x') \leq \frac{\delta}{4}$, then

$$|\bar{f}(x') - \bar{f}(x)| = |f(x') - f(x)| < \epsilon,$$

and this completes the proof.

Lemma 4.4.8 (Urysohn) Let (X, d) be a metric space, and A and B be disjoint closed subsets of X. Then there exists a continuous mapping $f : X \to [0, 1]$, where [0,1] is equipped with usual metric, such that $f(A) = \{0\}$ and $f(B) = \{1\}$.

Proof. The mapping f defined by

$$f(x) = \frac{d(x, A)}{(d(x, A) + d(x, B))}$$

for all x in X is continuous, and $f(A) = 0$ and $f(B) = 1$. Note that $d(x, A) + d(x, B) \neq 0$ by Theorem 2.9.6.

As an application of Gluing lemma 4.4.6, we show the continuity of a mapping in the following example.

Example 4.4.9 Let X be equal to complex plane $\mathbb{C}$ minus the non-positive real axis, that is, $X = \mathbb{C} - \{z \in \mathbb{C} : \text{Re } z \leq 0 \text{ and Im } z = 0\}$. Let $\alpha : X \to]-\pi, \pi[$ be a mapping defined

by $z = |z|e^{i\alpha(z)}$ for all z in X. We show that α is a continuous mapping on X. Let us define three open half-planes

$$H_1 = \{z \in \mathbb{C} : \operatorname{Re} z > 0\},$$
$$H_2 = \{z \in \mathbb{C} : \operatorname{Im} z > 0\},$$
$$H_3 = \{z \in \mathbb{C} : \operatorname{Im} z < 0\}.$$

Then $X = H_1 \cup H_2 \cup H_3$. Now let $z \in H_1$. Then $\operatorname{Re} z = |z| \cos\theta$ for some $\theta \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. The function $\sin x$ is continuous and monotonically increasing for $x \in]-\frac{\pi}{2}, \frac{\pi}{2}[$, so that its inverse $\sin^{-1} x$ is a continuous mapping from $]-1, 1[$ into $]-\frac{\pi}{2}, \frac{\pi}{2}[$. Since for any $z \in H_1$, $\frac{\operatorname{Im}(z)}{|z|} \in]-1, 1[$, the mapping $\alpha_1 : H_1 \to]-\frac{\pi}{2}, \frac{\pi}{2}[$ defined by $\alpha_1(z) = \sin^{-1}\left(\frac{\operatorname{Im} Z}{|z|}\right)$ is continuous. Similarly, the mapping $\alpha_2 : H_2 \to]0, \pi[$, and $\alpha_3 : H_3 \to]-\pi, 0[$ defined by

$$\alpha_2(z) = \cos^{-1}\left(\frac{\operatorname{Re} z}{|z|}\right) \quad \text{and} \quad \alpha_3(z) = \cos^{-1}\left(\frac{\operatorname{Re} z}{|z|}\right)$$

are continuous. Also α_1, α_2 and α_3 agree on their common domains, and $\alpha(z) = \alpha_i(z)$ for $z \in H_i$. Now, by Gluing Lemma 4.4.6, $\alpha(z)$ is continuous on X.

Let (X, d_1) and (Y, d_2) be metric spaces, and $A \subseteq X$. Let $f : A \to Y$ be a mapping. Then the existence of the limits $\lim_{\substack{y \to x \\ y \in A}} f(y)$ is sufficient to define a continuous mapping on the set of all such points x.

Theorem 4.4.10 Let (X, d_1) and (Y, d_2) be metric spaces, and $A \subseteq X$. Let $f : A \to Y$ be a mapping. Let $\widetilde{A}$ be the set of all points x in $\bar{A}$ such that $\lim_{\substack{y \to x \\ y \in A}} f(y)$ exists. Let $\bar{f} : \widetilde{A} \to Y$ be defined by

$$\bar{f}(x) = \lim_{\substack{y \to x \\ y \in A}} f(y).$$

Then $\bar{f}$ is continuous on $\widetilde{A}$.

Proof. Let x_0 be any point in $\widetilde{A}$, and $\bar{f}(x_0) = y_0$. Let $\epsilon > 0$ be a real number. Then there exists a real number $\delta > 0$ such that $f(S(x_0, \delta) \cap A) \subseteq S(y_0, \frac{\epsilon}{2})$. We show that $\bar{f}(S(x_0, \delta) \cap \widetilde{A}) \subseteq S(y_0, \epsilon)$. Let x be any point in $S(x_0, \delta) \cap \widetilde{A}$. Then there exists a real number $\delta' > 0$ such that $S(x, \delta') \subseteq S(x_0, \delta)$. Then $f(S(x, \delta') \cap A) \subseteq S(y_0, \frac{\epsilon}{2})$. Since $\bar{f}(x) \in \overline{[f(S(x, \delta') \cap A)]}$, it follows that $\bar{f}(x) \in \overline{S(y_0, \frac{\epsilon}{2})} \subseteq S(y_0, \epsilon)$. Therefore, $\bar{f}$ is continuous at x_0. Since x_0 is arbitrary, $\bar{f}$ is continuous on $\widetilde{A}$.

In the above theorem, $\widetilde{A} \cap A$ may or may not be empty. At any point $x \in A$ if $\lim_{\substack{y \to x \\ y \in A}} f(x)$ exists, then it is of course equal to $f(x)$, and $\bar{f}(x) = f(x)$. Thus, $\bar{f}$ and f coincide on $\widetilde{A} \cap A$. If $A \subseteq \widetilde{A}$, then f is continuous on A, and $\bar{f}$ is then the extension of f to $\widetilde{A}$.

Theorem 4.4.11 Let G be an open set in a metric space X. Then there exists a monotonically increasing sequence $\langle f_n \rangle$ of continuous mappings $f_n : X \to [0, 1], n \in \mathbb{Z}_+$, such that

$$\chi_G(x) = \lim_{n \to \infty} f_n(x) \text{ for all } x \text{ in } X$$

where χ_G is the characteristic mapping associated with G.

Proof. Since G is open, it is F_σ set (See Exercise 4.2(17)). Therefore, there exists a sequence $\langle F_n \rangle$ of closed sets such that $G = \bigcup F_n$. Since each F_n is a closed set contained

in G, $F_n \cap (X - G) = \emptyset$. Therefore, by Urysohn's Lemma 4.4.8, for each $n \in \mathbb{Z}$, there is a continuous mapping $f_n : X \to [0,1]$ such that $f_n(F_n) = 1$ and $f_n(X - G) = 0$.

Define the mappings $f'_n : X \to [0,1]$ inductively as follows:

$$f'_1 = f_1,$$

$$f'_n = f'_{n-1} \vee f_n \text{ for all } n \geq 2.$$

Then $\langle f'_n \rangle$ is a monotonically increasing sequence of continuous mappings. Therefore, for each x in X, the sequence $\langle f'_n(x) \rangle$ is monotonically increasing and bounded, and hence it is convergent. Denoting $f(x) = \lim_{n\to\infty} f'_n(x) = \sup\{f'_n(x) : n \in \mathbb{Z}_+\}$, it is easily seen that $f = \chi_G$.

The dual of the above theorem is stated below, and we consider it proved.

Corollary 4.4.12 Let H be a closed set in a metric space X. Then there exists a monotonically decreasing sequence $\langle f_n \rangle$ of continuous mappings $f_n : X \to [0,1]$, $n \in \mathbb{Z}_+$, such that

$$\chi_H(x) = \lim_{n\to\infty} f_n(x).$$

EXERCISE

1. Give an example of a continuous mapping $f : \mathbb{R} - \{0\} \to \mathbb{R}$ which cannot be extended to a continuous mapping from $\mathbb{R}$ into $\mathbb{R}$.

2. Let $n \to r_n$ be a bijective mapping from $\mathbb{N}$ onto the set A of all rational numbers in $[0,1]$. Let $f : [0,1] \to [0,1]$ be a function defined by

$$f(x) = \sum_{r_n < x} \frac{1}{2^n}.$$

The infinite sum extends only to those n such that $r_n < x$. Show that the restriction of f to the set B of all irrational numbers $x \in [0,1]$ is continuous, but it can not be extended to a continuous function on [0,1].

4.5 Homeomorphisms

The inverse of a bijective continuous mapping need not be continuous. For example, consider the discrete metric on real numbers; let $\mathbb{R}'$ be this metric space. Then the identity mapping $i : \mathbb{R}' \to \mathbb{R}$ is continuous, but i^{-1} is not. If a bijective continuous mapping f has the property that its inverse f^{-1} is also continuous, then f is called a homeomorphism. It turns out that a homeomorphism provides a one-to-one correspondence of open sets in domain and co-domain metric spaces, and this one-to-one correspondence is responsible for some simplification in the study of metric spaces. We begin by making the definition of a homeomorphism precise.

Definition 4.5.1 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping such that

(i) f is bijective,

(ii) f is continuous,

(iii) f^{-1} is continuous.

Then f is called a homeomorphism from X to Y, and X and Y are said to be homeomorphic, written $X \approx Y$.

Examples 4.5.2

1. The identity mapping from a metric space to itself is a homeomorphism.

2. The mapping $f : \mathbb{R} \to]-1, 1[$, where $]-1, 1[$ is equipped with the usual metric, defined by $f(x) = \frac{x}{(1+|x|)}$ for all x in $\mathbb{R}$ is a homeomorphism. Since $|f(x)| \leq 1$ for all $x \in \mathbb{R}$, the mapping f is well-defined. Now let y be any point in $]-1, 1[$. Then $f(x) = y$ implies that $x = \frac{y}{(1-|y|)}$ which is in $\mathbb{R}$. So f is surjective. Also $f(x_1) = f(x_2)$ implies that x_1 and x_2 are of the same sign, and $|x_1| = |x_2|$. Thus, f is injective. Since mappings $x \to 1$, $x \to |x|$, $x \to 1 + |x|$, $x \to x$ are continuous real valued mappings from $\mathbb{R}$ to $\mathbb{R}$, and $1 + |x| \neq 0$ for any x, the mapping $x \to \frac{x}{(1+|x|)}$ is continuous. Similarly, the inverse mapping $f^{-1}(x) = \frac{x}{(1-|x|)}$ is continuous. Thus, f is a homeomorphism, that is, $\mathbb{R} \approx]-1, 1[$.

3. Any two non-degenerate open intervals (An open interval $]a, b[$ in $\mathbb{R}$ is called non-degenerate if $a < b$) are homeomorphic to each other. Let $]a, b[$ and $]c, d[$ be two non-degenerate open intervals in $\mathbb{R}$ with usual metric on them. Let $f :]a, b[\to]c, d[$ be a mapping defined by $f(x) = Ax + B$, where A and B are constant so chosen that $f(a) = c$ and $f(b) = d$. That is, $f\left(\frac{(a+\lambda b)}{1+\lambda}\right) = \frac{(c+\lambda d)}{1+\lambda}$ for each real number λ such that $\lambda \neq -1$. Therefore, we choose A and B in such a way that the equality

$$A\frac{(a + \lambda b)}{(1 + \lambda)} + B = \frac{(c + \lambda d)}{(1 + \lambda)}$$

holds for each λ such that $\lambda \neq -1$. This gives $A = \frac{(d-c)}{(b-a)}$, and $B = \frac{(bc-ad)}{(b-a)}$. With these values of A and B, f is bijective. Since f is a linear mapping, it is continuous. The inverse mapping $f^{-1}(y) = \frac{(y-B)}{A}$ is also continuous. Thus, f is a homeomorphism, and $]a, b[\approx]c, d[$. From (2) in this example, it follows that $\mathbb{R} \approx]0, 1[$.

4. The mapping f in Example 4.2.19 is a bijective and continuous mapping. Since f^{-1} is not continuous, it is not a homeomorphism.

5. *A* metric space homeomorphic to a discrete metric space is sometimes called a discrete space even if its metric is different from that in Example 2.1.4.

   The following theorem are easy consequences of the definition of a homeomorphism.

Theorem 4.5.3 The inverse of a homeomorphism is a homeomorphism.

Theorem 4.5.4 The composition of two homeomorphisms is a homeomorphism.

Theorem 4.5.5 Every isometry is a homeomorphism. The converse of the above theorem is not true. The mapping $f : \mathbb{R} \to]-1, 1[$ defined by $f(x) = \frac{x}{(1+|x|)}$ is a homeomorphism, but not an isometry.

Theorem 4.5.6 Let (X, d_1) and (Y, d_2) be two metric spaces, and $f : X \to Y$ be a bijective mapping. Then the following statements are equivalent.

1. f is a homeomorphism.
2. A set G is open in X if and only if its image $f(G)$ is open in Y.
3. A set F is closed in X if and only if its image $f(F)$ is a closed set in Y.

Proof. (1) implies (2). Assume that f is a homeomorphism. Then f and f^{-1} are both continuous, and $(f^{-1})^{-1}(G) = f(G)$ for any subset G of X. If G is open in X, then $(f^{-1})^{-1}(G)$ is open in Y, that is, $f(G)$ is open in Y. If $G \subseteq X$ and $f(G)$ is open in Y, $f^{-1}(f(G))$ is open in X by the continuity of f. Since $G = f^{-1}(f(G))$, G is open in X.

(2) implies (3). Let F be a closed set in X. Then $X - F$ is open in X, and by (2), $f(X - F)$ is an open set in Y. Since $f(X - F) = f(X) - f(F)$, $f(F)$ is a closed set in Y. Conversely, if for any subset F of X, $f(F)$ is a closed set in Y, then $Y - f(F)$ is open in Y. Since $f(X - F) = Y - f(F)$, by (2), $X - F$ is open in X, and F is a closed set in X.

(3) implies (1). Let F be any closed set in X. Then $(f^{-1})^{-1}(F) = f(F)$ is a closed set in Y. Therefore, f^{-1} is continuous by Theorem 4.2.4. On the other hand, let G be an open set in Y. Then $Y - G$ is a closed set in Y, and $Y - G = f(X - f^{-1}(G))$. By (3), $X - f^{-1}(G)$ is a closed set in X so that $f^{-1}(G)$ is open in X. Thus, f is a homeomorphism.

Corollary 4.5.7 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a bijection. Then the following statements are equivalent.

1. f^{-1} is continuous.
2. $f(G)$ is open in Y for each open set G in X.
3. $f(F)$ is a closed set in Y for each closed set F in X.

Theorem 4.5.6 gives a one-to-one correspondence between the open sets in (X, d_1) and (Y, d_2). When $f : X \to Y$ is a homeomoprhism, then an open set G in X is mapped under f to an open set $f(G)$, and the open set $f(G)$ in Y is mapped by f^{-1} to the open set G. Thus, the open sets G and $f(G)$ are related via a homeomorphism (Fig. 4.2). Due to this one-to-one correspondence between open sets of X and Y, it turns out that any property P which can be defined entirely in terms of open sets in X, and which holds in X, also holds in Y. Such properties, as can be defined entirely in terms of open sets in a metric space, are known as topological properties. For example, the notions of neighbourhood, closed

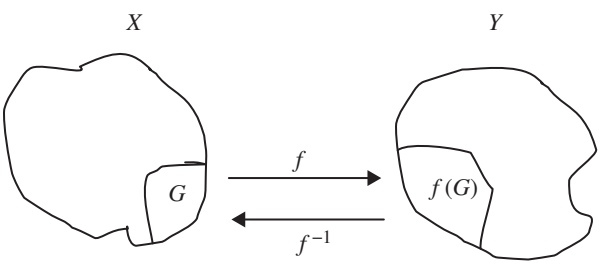


Fig. 4.2

set, adherent point, closure, interior, exterior, dense sets, boundary, continuous mapping etc. are topological properties. On the other hand, the notions of open spheres, diameter, bounded set, uniformly continuous mapping (defined later in this chapter) are some of the examples that are not topological properties. The topological properties remains invariant under homeomorphisms.

Theorem 4.5.8 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ and $g : Y \to X$ be continuous mappings such that $f \circ g = i_Y$ and $g \circ f = i_X$, where i_Y and i_X are identity mappings on Y and X, respectively. Then f is a homeomorphism, and, infact, $g = f^{-1}$.

Proof. It is easily seen that both f and g are bijective, and $g = f^{-1}$. Since f and g are both continuous, f is a homeomorphism.

Theorem 4.5.9 Let (X, d_1) and (Y, d_2) be metric spaces, and A be a non-empty subset of X. Let $f : X \to Y$ be a homeomorphism. Then the restriction $f_A : A \to f(A)$ is a homemorphism.

Proof. f_A is continuous by Theorem 4.2.18. Similarly $f^{-1}_{f(A)}$ is continuous on $f(A)$. Also f_A and $f^{-1}_{f(A)}$ satisfy the conditions of Theorem 4.5.8. Therefore f_A is a homeomorphism.

The following straightforward theorem gives another equivalent condition for two metrics to be equivalent.

Theorem 4.5.10 Let (X, d_1) and (X, d_2) be metric spaces. Then d_1 and d_2 are equivalent if and only if the identity mapping $i : (X, d_1) \to (X, d_2)$ is a homeomorphism.

The relation $X \approx Y$, that is, X is homeomorphic to Y in any set of metric spaces is an equivalence relation.

Theorem 4.5.11 Let $\mathcal{A}$ be any set of metric spaces. Then the relation $\approx$ on $\mathcal{A}$ is an equivalence relation.

Proof. Let X be a metric space in $\mathcal{A}$. Since $X \approx X$, the relation $\approx$ is reflexive. If X and Y are metric spaces in $\mathcal{A}$, and $X \approx Y$, then there exists a homeomorphism $f : X \to Y$. Since $f^{-1} : Y \to X$ is also a homeomorphism, $Y \approx X$. Since the composition of two homeomorphisms is a homeomorphism, the relation $\approx$ is transitive. Thus, the relation is an equivalence relation.

The above equivalence relation $\approx$ thus partitions a set of metric spaces into equivalence classes.

Theorem 4.5.12 Every separable metric space is homeomorphic to some subset of Fréchet metric space.

Proof. Let (X, d) be a separable metric space. Then there exists a countable dense set A in X. Since A is countable, we can arrange the elements in A in the form of a sequence $\langle x_n \rangle$. Let (Y, σ) be the Fréchet metric space. Let $f : X \to Y$ be the mapping defined by $f(x) = \langle d(x, x_1), d(x, x_2), \ldots \rangle$. This is clearly a well defined mapping of X onto the subset $f(X)$ of Y. We show that f is a homeomorphism from X onto $f(X)$. Let x and x' be any two points in X. Then $f(x) = f(x')$ implies that $d(x, x_n) = d(x', x_n)$ for all $n \in \mathbb{Z}_+$. Let $\epsilon > 0$ be a real number. Since A is dense in X, there exists an $n \in \mathbb{Z}_+$ such that $d(x, x_n) < \frac{\epsilon}{2}$. Then $d(x', x_n) < \frac{\epsilon}{2}$. Therefore, $d(x, x') < \epsilon$. Since ϵ is arbitrary, $d(x, x') = 0$, that is, $x = x'$. To prove that f is continuous, assume that $\langle y_n \rangle$ is a sequence in X which converges to $y \in X$. Then the sequence $\langle f(y_n) \rangle$ converges in $f(X)$ to $f(y) \in f(X)$ if and only if for each $j \in \mathbb{Z}_+$, the sequence in jth co-ordinates $\langle d(y_n, x_j) \rangle$ of $\langle f(y_n) \rangle$ converges in $\mathbb{R}$ to the jth co-ordinate $d(y, x_j)$ of $f(y)$. But this was shown in Theorem 3.3.3. Finally, we

show that f^{-1} is continuous. Let $\langle y_n \rangle$ be a sequence in $f(X)$ converging to y in $f(X)$. Let $\epsilon > 0$ be a real number. Let $f^{-1}(y_n) = z_n \in X$, $f^{-1}(y) = z \in X$. Then $f(z_n) = y_n$ and $f(z) = y$. We need to show that the sequence $\langle f^{-1}(y_n) \rangle$ converges in X to $f^{-1}(y)$, that is, the sequence $\langle z_n \rangle$ converges in X to z. Since A is dense in X, there exists a $k \in \mathbb{Z}_+$ such that $d(z, x_k) < \frac{\epsilon}{2}$. Now $\langle y_n \rangle$ converges to y in $f(X)$ if and only if the sequence in jth co-ordinates $\langle d(z_n, x_j) \rangle$ in the sequence $\langle y_n \rangle$ converges in $\mathbb{R}$ to jth co-ordinate $d(z, x_j)$ of y. Therefore,

$$d(z_n, z) \leq d(z_n, x_k) + d(x_k, z).$$

Taking limits as n tends to infinity, we have

$$\lim_{n \to \infty} d(z_n, z) \leq \lim_{n \to \infty} d(z_n, x_k) + d(x_k, z) < \epsilon.$$

Since ϵ is arbitrary, it follows that $\langle z_n \rangle$ converges to z.

Example 4.5.13 Let S^n be the unit sphere defined by

$$S^n = \left\{ (x_1, x_2 \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : \sum_{i=l}^{n+1} x_i^2 = 1 \right\}.$$

Let $p = (0, 0, \ldots, 0, 1)$ be the north pole of S^n. We show that the subspace $S^n - \{p\}$ of $\mathbb{R}^{n+1}$ is homeomorphic to $\mathbb{R}^n$. Now the set of points of $\mathbb{R}^{n+1}$ having their $(n+1)^{\text{th}}$ co-ordinate equal to zero with the subspace metric is clearly homeomorphic to $\mathbb{R}^n$. We define a mapping $f : S^n - \{p\} \to \mathbb{R}^n$, called stereographic projection, as follows: for any x in $S^n - \{p\}$, $f(x)$ is the point of intersection with $\mathbb{R}^n$ of the straightline determined by x and p (see figure in Example 2.1.17 for $n = 2$). f is clearly a bijecive mapping. If G is an open set in $\mathbb{R}^n$, then $f^{-1}(G)$ is the set of points x in $S^n - \{p\}$ such that the straightline joining p and x pass through a point in G. One readily checks that $f^{-1}(G)$ is an open set in $S^n - \{p\}$. Thus, f is continuous. A similar argument will show that f^{-1} is a continuous mapping. Therefore, f is a homeomorphism.

EXERCISE

1. Let (X, d_1) and (Y, d_2) be homeomorphic metric spaces. Show that (X, d_1) is discrete if and only if (Y, d_2) is discrete.

2. Let X and Y be non-empty equinumerous sets. Let d and d' be discrete metrics associated with X and Y, respectively. Show that (X, d) and (Y, d') are homeomorphic.

3. Let $f : (X, d_1) \to (Y, d_2)$ be a homeomorphism, $A \subseteq X$, and $x \in A$. Prove that

   (a) x is a limit point of A if and only if $f(x)$ is a limit point of $f(A)$.

   (b) x is an interior point of A if and only if $f(x)$ is an interior point of $f(A)$.

4. Let $f : (X, d_1) \to (Y, d_2)$ be a homeomorphism, and A be a subset of X such that $A \cap D(A) = \emptyset$, where $D(A)$ is the set of all limit points of A in (X, d). Show that $f(A) \cap D(f(A))$ is empty.

5. Let X, Y and Z be three sets, and $f : X \to Y$, $g : Y \to Z$ be mappings such that $g \circ f$ is bijective. If either g is injective, or f is surjective , show that f and g are both bijective. Let d_1, d_2 and d_3 be metrics on X, Y and Z respectively. If f and g are continuous, and $g \circ f$ is a homeomorphism, prove that f and g both are homeomorphisms.

6. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. Let k_1 and k_2 be the positive real numbers such that

$$k_1 d_1(x, y) \leq d_2(f(x), f(y)) \leq k_2 d_1(x, y) \text{ for all } x, y \text{ in} X.$$

   Then X and Y are homeomorphic.

7. Let (X, d) and (Y, d') be metric spaces, and A be a dense set in X. Let $f : X \to Y$ be a continuous mapping. Assume that the restriction $f_A : A \to Y$ is a homeomorphism. Prove that $f(X - A) \subseteq Y - f(A)$.

8. Let A be a dense subset of each of the metric spaces (X, d_1) and (Y, d_2). Let the identity mapping $i : A \to A$ be extendable to a continuous mapping $f : X \to Y$ and also to a continuous mapping $g : Y \to X$. Prove that f and g are homeomorphisms, and that $f = g^{-1}$.

9. Let X and Y be metric spaces, and $f : X \to Y$ be a continuous mapping. Let $\{G_i\}_{i \in \Lambda}$ be an open covering of Y, that is, $\bigcup_{i \in \Lambda} G_i = Y$, and each G_i is open in Y. Show that if for each $i \in \Lambda$, the restriction of f to $f^{-1}(G_i)$ is a homeomorphism of the subspace $f^{-1}(G_i)$ of X onto the subspace G_i of Y, then f is a homeomorphism.

10. Give examples of two metric spaces X and Y, a continuous mapping $f : X \to Y$ which is not injective, and for an open covering $\{G_i\}_{i \in \Lambda}$ such that the restriction of f to each G_i is a homeomorphism of the subspace G_i of X onto the subspace $f(G_i)$ of Y (one can take both X and Y discrete metric spaces)

11. (Brouwer (1913)) Prove that for any two countable dense subsets A and B of the real line $\mathbb{R}$, there exists a homeomorphism $f : \mathbb{R} \to \mathbb{R}$ such that $f(A) = B$.

12. Prove that if A_1, A_2, B_1 and B_2 are countable dense subsets of $\mathbb{R}$ satisfying the condition $A_1 \cap A_2 = \emptyset = B_1 \cap B_2$, then there exists a homeomorphism $f : \mathbb{R} \to \mathbb{R}$ such that $f(A_i) = B_i$ for $i = 1, 2$.

13. Show that for any two countable dense subsets A and B of the space P of irrational numbers (with usual metric), there exists a homeomorphism $f : P \to P$ such that $f(A) = B$.

14. Show that for any two countable dense subsets A and B of the Cantor set C, there exists a homeomorphism $f : C \to C$ such that $f(A) = B$.

15. Let (X, d) and (Y, d') be homeomorphic. Prove that (X, d) is separable if and only if (Y, d') is separable.

16. Prove that the set of rational numbers is not homeomorphic to the set $\{\frac{1}{n} : n \in \mathbb{Z}_+\} \cup \{0\}$. Show that the set of all rational numbers $\mathbb{Q}$ is homeomorphic to the set $\mathbb{Q}_+$ of positive rational numbers. (All sets are equipped with usual metric in this example.)

17. Show that the closed unit sphere $S[0,1] = \left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq 1\right\}$ is homeomorphic to the cube $[-1,1]^n \subseteq \mathbb{R}^n$.

18. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a homeomorphism. Prove that the mapping ρ defined by $\rho(x, y) = d_2(f(x), f(y))$ is a metric on X which is equivalent to d_1.

19. Show that the subspaces $]0,1]$ and $[1,\infty[$ of the real line $\mathbb{R}$ are homeomorphic. Also show that any two open spheres in $\mathbb{R}^n$ are homeomorphic.

20. Show that a circle and an ellipse in $\mathbb{R}^2$ are homeomorphic.

21. Show that the real line $\mathbb{R}$ is homeomorphic with the parabola $y = x^2$ in $\mathbb{R}^2$.

22. Show that the function $f(x) = \frac{e^x}{1+e^x}$ is a homeomorphism from the real line to the open interval $]0,1[$.

23. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous mapping. We define the graph of f as $G_f : \mathbb{R} \to \mathbb{R}^2$ given by $G_f(x) = (x, f(x))$. Show that G_f is a continuous mapping, and that its image taken with the subspace metric from $\mathbb{R}^2$ is homeomorphic with $\mathbb{R}$.

4.6 Open Mappings and Closed Mappings

If $f : (X, d_1) \to (Y, d_2)$ is a continuous mapping, and A is an open set in X, then it is not necessary that $f(A)$ is an open set in Y. For example, let (X, d) be a discrete metric space, then any mapping $f : X \to \mathbb{R}$ is continuous on X. If x is any point in X, then the set $\{x\}$ is open in X but the set $f(\{x\})$ is the singleton $\{f(x)\}$ which is not open in the real line $\mathbb{R}$. Similarly, the image of a closed set under a continuous mapping need not be a closed set. This is the motivation for the next definition.

Definition 4.6.1 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. Then f is said to an open mapping if for each open set A in X, the set $f(A)$ is open in Y. f is said to be a closed mapping if for each closed set F in X the image set $f(F)$ is a closed set in Y.

An open mapping need not be continuous. The identity mapping $i : (X, d_1) \to (X, d_2)$ is open if and only if the metric d_2 is finer than the metric d_1. If we equip $\mathbb{R}$ with discrete metric d_1, then the identity mapping from the real line onto the discrete metric space $(\mathbb{R}, d_1)$ is open, but not continuous.

Theorem 4.6.2 Let X and Y be metric spaces, and $f : X \to Y$ be a mapping. Then the following statements hold.

1. If f is open, then for each subset A of Y, and any closed set F containing $f^{-1}(A)$, there exists a closed set F^* in Y containing A, and is such that $f^{-1}(F^*) \subseteq F$.

2. If f is closed, then for each subset A of Y, and any open set G containing $f^{-1}(A)$, there exists an open set G^* in Y containing A, and is such that $f^{-1}(G^*) \subseteq G$.

Proof. (1) Let $F^* = Y - f(X - F)$. Then F^* is a closed set, $A \subseteq F^*$, and

$$f^{-1}(F^*) = X - f^{-1}(f(X - F)) \subseteq X - (X - F) = F.$$

The proof of (2) is similar.

Theorem 4.6.3 Let X and Y be metric spaces, and $f : X \to Y$ be a mapping. Then the following statements are equivalent.

1. f is an open mapping.
2. $f(A^0) \subseteq (f(A))^0$ for each subset A of X.
3. For each $x \in X$ and a neighbourhood N of x, there exists a neighbourhood M in Y of $f(x)$ such that $M \subseteq f(N)$.

Proof. (1) implies (2). Since $A^0 \subseteq A$, $f(A^0) \subseteq f(A)$. By our assumption, $f(A^0)$ is open set. Therefore, $f(A^0) = (f(A^0))^0 \subseteq (f(A))^0$.

(2) implies (3). Let N be a neighbourhood of x in X. Then there exists an open sphere $S(x, \epsilon)$ contained in N. By our assumption, $f(S(x, \epsilon)^0) \subseteq (f(S(x, \epsilon)))^0$. Therefore, $f(x) \in f(S(x, \epsilon)) \subseteq (f(S(x, \epsilon)))^0 \subseteq f(N)$. Now take $M = (f(S(x, \epsilon)))^0$.

(3) implies (1). Let G be an open set in X. Let y be a point in $f(G)$. By our assumption, there exists an open neighbourhood M_y in Y of y such that $M_y \subseteq f(G)$. Then $f(G) = \bigcup_{y \in f(G)} M_y$, and $f(G)$ is an open set in Y.

Theorem 4.6.4 Let (X, d_1) and (Y, d_2) be metric spaces. A mapping $f : X \to Y$ is closed if and only if $\overline{f(A)} \subseteq f(\bar{A})$ for every subset A of X.

Proof. Let f be a closed mapping. Then $f(\bar{A})$ is a closed set in Y. Since $f(A) \subseteq f(\bar{A})$, $\overline{f(A)} \subseteq \overline{(\bar{A})} = f(\bar{A})$. Conversely, assume that $\overline{f(A)} \subseteq f(\bar{A})$ holds for each subset A of X . Let A be a closed set in X. Then $\overline{f(A)} \subseteq f(\bar{A})$. Thus, $\overline{f(A)} = f(A)$.

The following theorem obtains further equivalent conditions for a homeomorphism.

Theorem 4.6.5 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping. The following statements are equivalent.

1. f is a homeomorphism.
2. f is continuous and open.
3. f is continuous and closed.
4. $f(\bar{A}) = \overline{f(A)}$ for each subset A of X.

Proof. The equivalences of the statements (1), (2) and (3) are already proved. We prove here the equivalence of (3) and (4). Let f be continuous and closed. By the continuity of f, Theorem 4.2.10 gives that $f(\bar{A}) \subseteq \overline{f(A)}$. Since f is closed, Theorem 4.6.4 gives that $\overline{f(A)} \subseteq f(\bar{A})$, and so (3) implies (4). Conversely, the continuity of f follows from Theorem 4.2.10, and it is obvious that f is closed.

A natural question arises here: is a closed mapping an open mapping? There is also the converse question : is an open mapping a closed mapping ? The answer is provided in the following example.

Example 4.6.6 Consider the mapping $f : \mathbb{R} \to S^1$, where S^1 is the unit circle in $\mathbb{R}^2$, defined by $f(t) = (\cos \pi t, \sin \pi t)$. Geometrically, this mapping wraps the real line around the circle infinite number of times. Clearly, this mapping is open. The set of point $\{2n + \frac{1}{n} : n \in \mathbb{Z}_+\}$ is a closed set in $\mathbb{R}$, for it has no limit point in $\mathbb{R}$; however, its image under f is the set $\{\cos(2\pi n + \frac{\pi}{n}), \sin(2\pi n + \frac{\pi}{n}) : n \in \mathbb{Z}_+\}$ which has $(1, 0)$ as its limit point, and therefore not a closed set. To settle the second question, consider the mapping $g : [-2, 2] \to [-2, 2]$ defined by $g(t) = \frac{(t^5+7t^3-26t)}{18}$. Its graph is shown in Fig. 4.3

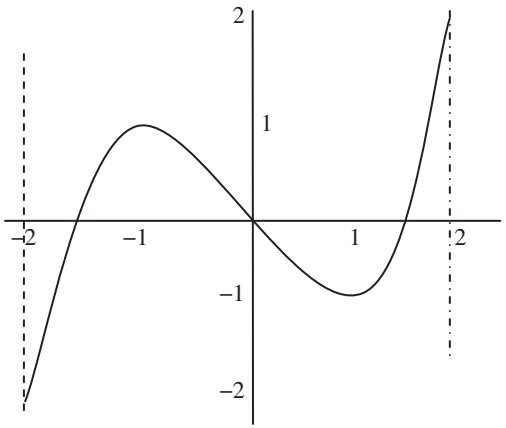


Fig. 4.3

The mapping g is a closed mapping. The mapping g is not open. The image of the open set $\{t : 0 < t \leq 2\}$ in $[-2, 2]$ under g is the closed set $[-1, 2]$.

EXERCISE

1. Show that the mapping $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x$ is continuous, but it is not closed.

2. Let X, Y and Z be metric spaces, and $f : X \to Y$ and $g : Y \to Z$ be mappings. Then prove the following statements.

   (a) If f is continuous and onto, and $g \circ f$ is open, then g is open.

   (b) If g is continuous and injective, and $g \circ f$ is open, then f is open.

   (c) If we replace in (a) and (b) "open" by "closed", are the resulting statements true?

3. Let Y be a subspace of a metric space X. Check whether the inclusion mapping $i : Y \to X$ is open or closed.

4. Let X and Y be metric spaces. A bijective mapping $f : X \to Y$ is open if and only if f is closed.

4.7 Uniform Convergence

In this section, we discuss the convergence of sequences of mappings.

Definition 4.7.1 Let X be a set, and (Y, d) be a metric space. Then a sequence $\langle f_n \rangle$ of mappings $f_n : X \to Y$ is said to converge uniformly on X to a mapping $f : X \to Y$ if for each real number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that

$$d(f_n(x), f(x)) < \epsilon \text{ for all } n \geq m \text{ and all } x \in X.$$

We have the following theorem on the uniform convergence of continuous mappings.

Theorem 4.7.2 Let (X, d_1) and (Y, d_2) be metric spaces, and $\langle f_n \rangle$ be a sequence of continuous mappings $f_n : X \to Y$. If $\langle f_n \rangle$ converges uniformly to a mapping $f : X \to Y$, then f is continuous.

Proof. Let x_0 be a point in X, and $\epsilon > 0$ be a real number. Since $\langle f_n \rangle$ uniformly converges to f, there exists an $m \in \mathbb{Z}_+$ such that

$$d_2(f_n(x), f(x)) < \frac{\epsilon}{3} \quad \text{for all } n \geq m \text{ and all } x \in X.$$

In particular,

$$d_2(f_m(x), f(x)) < \frac{\epsilon}{3} \quad \text{for all } x \in X.$$

Since f_m is continuous, there exists a real number $\delta > 0$ such that

$$d_1(x, x_0) < \delta \quad \text{implies that} \quad d_2(f_m(x), f_m(x_0)) < \frac{\epsilon}{3}.$$

Therefore, $d_1(x, x_0) < \delta$ implies that

$$\begin{aligned} d_2(f(x), f(x_0)) &\leq d_2(f(x), f_m(x)) + d_2(f_m(x), f_m(x_0)) + d_2(f_m(x_0), f(x_0)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

and the proof is complete.

Now let us consider the case when $Y = \mathbb{C}$, the complex plane, in the above theorem. Let $f_n : X \to \mathbb{C}$, and $s_n(x) = f_1(x) + f_2(x) + \cdots + f_n(x)$. For $x \in X$, we say that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ if $\lim_{n\to\infty} s_n(x) = f(x)$. The series $\sum_{n=1}^{\infty} f_n$ is said to be uniformly convergent to f if the sequence $\langle s_n \rangle$ converges uniformly to f.

Theorem 4.7.3 (Weirstrass M-Test) For each $n \in \mathbb{Z}_+$, let $f_n : X \to \mathbb{C}$ be a mapping, and M_n be a real number such that $|f_n(x)| \leq M_n$ for all x in X. If the series $\sum_{n=1}^{\infty} M_n$ is convergent in the real line $\mathbb{R}$, then $\sum_{n=1}^{\infty} f_n$ is uniformly convergent.

Proof. Let $s_n(x) = f_1(x) + f_2(x) + \cdots + f_n(x)$. For $n > m$, we have

$$|s_n(x) - s_m(x)| = |f_{m+1}(x) + f_{m+2}(x) + \cdots + f_n(x)| \leq \sum_{k=m+1}^{n} M_k$$

for each x in X. Since the series $\sum_{n=1}^{\infty} M_n$ is convergent in the real line $\mathbb{R}$, it follows that $\langle s_n(x) \rangle$ is a Cauchy sequence in $\mathbb{C}$ for each $x \in X$, and so converges to a point ξ_x in $\mathbb{C}$.

We define a mapping $f : X \to \mathbb{C}$ by $f(x) = \xi_x$. Now let $\epsilon > 0$ be a real number. Since $\lim_{n\to\infty} s_n(x) = f(x)$, $f(x) = \sum_{n=1}^{\infty} f_n(x)$, and

$$|s_n(x) - f(x)| = \left| \sum_{k=n+1}^{\infty} f_k(x) \right| \leq \sum_{k=n+1}^{\infty} |f_k(x)| \leq \sum_{k=n+1}^{\infty} M_k.$$

Since $\sum_{k=1}^{\infty} M_k$ is convergent, there exists an $n_0 \in \mathbb{Z}_+$ such that

$$\sum_{k=n+1}^{\infty} M_k < \epsilon \text{ for all } n \geq n_0.$$

This gives that $|s_n(x) - f(x)| < \epsilon$ for all $n \geq n_0$ and $x \in X$.

Example 4.7.4 Let (X, d) be a metric space, and Y be a set. In Example 2.4.6, we have shown that $\mathbb{B}(Y, X)$ is a metric space with metric ρ given by

$$\rho(\phi, \psi) = \sup\{d(\phi(y), \psi(y)) : y \in Y\}.$$

A sequence $\langle \phi_n \rangle$ in $\mathbb{B}(Y, X)$ converges to a limit ϕ in $\mathbb{B}(Y, X)$ if and only if for each real number $\epsilon > 0$, there exists an $n_0 \in \mathbb{Z}_+$ such that

$$d(\phi_n(y), \phi(y)) < \epsilon \text{ for all } n \geq n_0 \text{ and all } y \in Y.$$

In other words, a sequence in $\mathbb{B}(Y, X)$ converges to a limit in $\mathbb{B}(Y, X)$ if and only if it converges to that limit uniformly. For this reason the metric ρ on $\mathbb{B}(Y, X)$ is known as the metric of uniform convergence on Y.

The reader may note that the metric space $\mathcal{C}[a, b]$ is a subspace of the metric space $\mathbb{B}([a, b], \mathbb{R})$, and hence the convergence of a sequence in $\mathcal{C}[a, b]$ is equivalent to the uniform convergence of that sequence on $[a, b]$.

EXERCISE

1. Let $\mathcal{C}(\mathbb{R})$ be the set of all continuous functions from the real line $\mathbb{R}$ into the real line $\mathbb{R}$. For each $i \in \mathbb{Z}_+$, define

   $$\rho_i(f, g) = \sup_{|t| \leq i} \{|f(t) - g(t)|\}$$

   and

   $$\rho(f, g) = \sum_{i=1}^{\infty} \frac{\rho_i(f, g)}{[1 + \rho_i(f, g)]}.$$

   Show that $(\mathcal{C}(\mathbb{R}), \rho)$ is a metric space. A sequence $\langle f_n \rangle$ in $\mathcal{C}(\mathbb{R})$ converges to f in $\mathcal{C}(\mathbb{R})$ if and only if $\langle f_n \rangle$ converges uniformly to f over $[-i, i]$ for all $i > 0$.

4.8 Uniformly Continuous Mappings

In the notion of continuity of a mapping at a point x_0 (Definition 4.1.2), the real number δ depends on both the real number ϵ and the point x_0. If the point x_0 is changed, the real number δ may be different for the same value of ϵ. For example, consider the mapping $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. It is easily seen that $f(x)$ is continuous. Let $\epsilon = 1$, and $x_0 = 0$. Then

$$|f(x) - f(x_0)| = |x^2 - 0| < 1 \text{ if } |x - 0| < 1.$$

So if we take $\delta = 1$, then

$$|x - x_0| < \delta \text{ implies that } |f(x) - f(x_0)| < \epsilon.$$

Now, let us take $x_0 = 10$, and let $x = 10 + \frac{1}{2}$. Then $|x - x_0| < \delta$, but

$$|f(x) - f(x_0)| = |f(10 + \frac{1}{2}) - f(10)| = |(10 + \frac{1}{2})^2 - 10^2| > 1.$$

Therefore, at $x_0 = 10$, and for $\epsilon = 1$, the required δ in the definition of continuity can not be 1, and some other value of δ must be found. This new value of δ will of course exist, and will be less than 1. In fact, it is not possible to find a value of δ that will meet the requirement of continuity simultaneously at every point of $\mathbb{R}$ for the value 1 of ϵ. However, there are mappings for which a real number δ can be found for a given value of ϵ such that δ meet the requirement of continuity at every point of the domain of the mapping. Such mappings are known as uniformly continuous.

Definition 4.8.1 Let (X, d_1) and (Y, d_2) be metric spaces. A mapping $f : X \to Y$ is said to be uniformly continuous on X if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ depending only on ϵ such that

$$d_1(x, x') < \delta \text{ implies that } d_2(f(x), f(x')) < \epsilon \text{ for all } x, x' \text{ in } X.$$

If in the above definition, (X, d_1) is a discrete metric space, then any mapping $f : X \to Y$ is uniformly continuous. In general, a uniformly continuous mapping is continuous.

Theorem 4.8.2 Every uniformly continuous mapping is continuous.

Proof. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a uniformly continuous mapping. Let $\epsilon > 0$ be a real number. Then there exists a real number $\delta > 0$ such that

$$d_1(x, x') < \delta \text{ implies that } d_2(f(x), f(x')) < \epsilon \text{ for all } x, x' \in X.$$

Now take $x' = x_0$, and fix it. Then

$$d_1(x, x_0) < \delta \text{ implies that } d_2(f(x), f(x_0)) < \epsilon \text{ for all } x \in X.$$

But this means that f is continuous at x_0. Since x_0 is arbitrary point in X, f is continuous on X.

The converse of the above theorem is not true. Let $X = \mathbb{R} - \{0\}$, and equip X with usual metric. Let $f : X \to \mathbb{R}^2$ be the mapping defined by $f(x) = (x, \frac{1}{x})$. Then f is a continuous mapping, but f is not uniformly continuous. Let $\epsilon > 0$ be any real number. Then for any real number $\delta > 0$,

$$d(f(x + \frac{\delta}{2}), f(x)) = \frac{\delta}{2}\sqrt{1 + \frac{1}{[(x + \frac{\delta}{2})x]^2}}$$

can be made arbitrary large, and thus is not less than ϵ for all $x \in X$. Therefore, f is not uniformly continuous on X.

Theorem 4.8.3 Let (X, d) be a metric space, and A be a non-empty subset of X. Then the mapping $f : X \to \mathbb{R}$ defined by $f(x) = d(x, A)$ is uniformly continuous on X.

Proof. The proof follows from Lemma 2.4.5.

Theorem 4.8.4 Let (X, d) be a metric space. Then the metric $d : X \times X \to \mathbb{R}$, where $X \times X$ is equipped with the product metric d^*, is uniformly continuous mapping on X.

Proof. Let $\epsilon > 0$ be a real number. By Lemma 3.1.6,

$$|d(x_1, x_2) - d(y_1, y_2)| \leq d(x_1, y_1) + d(x_2, y_2).$$

Let us choose $\delta = \frac{\epsilon}{2}$. Then

$$d^*((x_1, x_2), (y_1, y_2)) < \delta \text{ implies that } \max\{d(x_1, y_1), d(x_2, y_2)\} < \delta.$$

Therefore,

$$d^*((x_1, x_2), (y_1, y_2)) < \delta \text{ implies that } |d(x_1, x_2) - d(y_1, y_2)| < \epsilon.$$

Theorem 4.8.5 Let (X, d_1), (Y, d_2) and (Z, d_3) be metric spaces. Let $f : X \to Y$ and $g : Y \to Z$ be uniformly continuous mappings. Then $g \circ f$ is uniformly continuous on X.

Proof. Let $\epsilon > 0$ be a real number. Then there exist real numbers $\delta > 0$, and $\eta > 0$ such that

$$d_2(y, y') < \eta \quad \text{implies that } d_3(g(y), g(y')) < \epsilon$$

and

$$d_1(x, x') < \delta \quad \text{implies that } d_2(f(x), f(x')) < \eta.$$

Then $d_1(x, x') < \delta$ implies that $d_3((g \circ f)(x), (g \circ f)(x')) < \epsilon$. Therefore, $g \circ f$ is uniformly continuous.

Example 4.8.6 Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a mapping such that

$$d_2(f(x), f(y)) \leq M d_1(x, y) \text{ for all } x, y \text{ in } X$$

where $M > 0$ is a real number. Then f is called a Lipschitz mapping. A Lipschitz mapping is uniformly continuous. For a real number $\epsilon > 0$, choose $\delta = \frac{\epsilon}{M}$.

Theorem 4.8.7 Every isometry is uniformly continuous.

Proof. The proof is obvious from the definition of an isometry.

Definition 4.8.8 Let (X, d_1) and (X, d_2) be metric spaces. d_1 and d_2 are said to be uniformly equivalent if the identity mapping $i : (X, d_1) \to (X, d_2)$ and its inverse $i^{-1} : (X, d_2) \to (X, d_1)$ are uniformly continuous mappings.

It follows from the above definition that if d_1 and d_2 are uniformly equivalent, then the identity mapping i is a homeomorphism, and, by Theorem 4.5.10, the metrics d_1 and d_2 are equivalent.

The following theorem gives a sufficient condition for two metrics to be uniformly equivalent.

Theorem 4.8.9 Let (X, d_1) and (X, d_2) be two metric spaces. If there exist positive real numbers k_1 and k_2 such that

$$k_1 d_1(x, y) \leq d_2(x, y) \leq k_2 d_1(x, y) \text{ for all } x, y \text{ in } X,$$

then $d_1$and d_2 are uniformly equivalent.

Proof. Let $i : (X, d_1) \to (X, d_2)$ be the identity mapping. Let $\epsilon > 0$ be a real number. Let us take $\delta = \frac{\epsilon}{k_2}$. Then, by the second inequality of our assumption,

$$d_1(x, y) < \delta \text{ implies that } d_2(i(x), i(y)) = d_2(x, y) < \epsilon.$$

Thus, i is uniformly continuous. Similarly, i^{-1} is uniformly continuous.

The conditions of the above theorem are not necessary when d_1 and d_2 are uniformly equivalent.

Example 4.8.10 Let (X, d_1) be an unbounded metric space. Then the mapping $d_2 : X \times X \to \mathbb{R}$ defined by $d_2(x, y) = \frac{d_1(x,y)}{(1+d_1(x,y))}$ for all $x, y \in X$ is a metric on X. We show that d_1 and d_2 are uniformly equivalent. (We have already shown them to be equivalent.) Also note that for this part we do not need the fact that d_1 is unbounded. Since $d_2(x, y) \leq d_1(x, y)$, the uniform continuity of the identify mapping $i : (X, d_1) \to (X, d_2)$ is clear. Thus, it remains to show that $i^{-1} : (X, d_2) \to (X, d_1)$ is uniformly continuous. Let $\epsilon > 0$ be a real number. Then

$$d_1(x, y) < \epsilon \quad \text{whenever} \quad \frac{d_1(x, y)}{(1 + d_1(x, y))} < \frac{\epsilon}{(1 + \epsilon)}.$$

Let $\delta = \frac{\epsilon}{(1+\epsilon)}$. Then

$$d_2(x, y) < \delta \quad \text{implies that} \quad d_1(i^{-1}(x), i^{-1}(y)) = d_1(x, y) < \epsilon.$$

Thus, i^{-1} is uniformly continuous. On the other hand, since d_1 is a unbounded metric, there is no constant $k_1 > 0$ such that $k_1 d_1(x, y) \leq d_2(x, y)$ for all x, y in X.

Theorem 4.8.11 Let (X, d_1) be a metric space. Then there exists a metric d_2 on X uniformly equivalent to d_1 such that (X, d_2) is bounded.

Proof. The metric d_2 of the previous example satisfies the required property.

Finally, it is noted that the relation of two metrics being uniformly equivalent is also an equivalence relation on any set of metric spaces $\mathcal{A} = \{(X, d_i)\}_{i \in \Lambda}$, and thus partitions $\mathcal{A}$ into equivalence classes.

We close this section by showing that the image under uniformly continuous mapping of a totally bounded set is totally bounded.

Theorem 4.8.12 *A* uniformly continuous image of a totally bounded subset of a metric space is totally bounded.

Proof. Let (X, d_1) and (Y, d_2) be metric spaces, and A be a totally bounded subset of X. Let $f : X \to Y$ be a uniformly continuous mapping. Let $\epsilon > 0$ be a real number. Since f is uniformly continuous, there exists a real number $\delta > 0$ such that

$$d_1(x, y) < \delta \quad \text{implies that} \quad d_2(f(x), f(y)) < \epsilon.$$

Since A is totally bounded, there exists a finite δ-net $\{x_1, x_2, \ldots, x_n\}$ for A in A. We show that the set $\{f(x_1), f(x_2), \ldots, f(x_n)\}$ is a finite ϵ-net for $f(A)$. Let y be any point

in $f(A)$. Then there exists a point x in A such that $f(x) = y$. Then there exists a point x_i in $\{x_1, x_2, \ldots, x_n\}$ such that $d_1(x, x_i) < \delta$. Therefore, $d_2(f(x), f(x_i)) < \epsilon$, that is, $d_2(y, f(x_i)) < \epsilon$.

Since every isometry is uniformly continuous, the total boundedness is preserved under isometries.

The continuous image of a totally bounded set may not be totally bounded. For example, the function $f :]0, 1[\to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ is continuous. The interval $]0, 1[$ is totally bounded, but its image $f(]0, 1[)$ is not bounded.

Theorem 4.8.13 Let V be a real (respectively, complex) normed space. Then the mapping $(x, y) \to x + y$ is uniformly continuous in $V \times V$. The mapping $(\lambda, x) \to \lambda x$ is continuous in $\mathbb{R} \times V$ (respectively, $\mathbb{C} \times V$). The mapping $x \to \lambda x$ is uniformly continuous in V.

Proof. The proof of continuity of $(\lambda, x) \to \lambda x$ at a point (λ_0, x_0) follows from the following inequality

$$\begin{aligned}\|\lambda x - \lambda_0 x_0\| &= \|\lambda_0(x - x_0) + (\lambda - \lambda_0)x_0 + (\lambda - \lambda_0)(x - x_0)\| \\ &\leq |\lambda_0| \, \|x - x_0\| + |\lambda - \lambda_0| \, \|x_0\| + |\lambda - \lambda_0| \, \|x - x_0\|.\end{aligned}$$

EXERCISE

1. Define a mapping $d : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ by $d(m, n) = |\frac{1}{m} - \frac{1}{n}|$. Show that $(\mathbb{N}, d)$ is metric space, and d is equivalent to the standard discrete metric on $\mathbb{N}$, but that d is not uniformly equivalent to the standard discrete metric on $\mathbb{N}$.

2. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a continuous mapping. Then $d(x, y) = d_1(x, y) + d_2(f(x), f(y))$ is equivalent to the metric d_1, and the mapping $f' : (X, d) \to (Y, d_2)$ defined by $f'(x) = f(x)$ for all x in X is uniformly continuous.

3. If (X, d) is totally bounded, and ρ is a metric on X which is uniformly equivalent to d, then (X, ρ) is totally bounded.

4. Let (X, d) be a metric space, and x_0 be a point in X. Show that the mapping $f : X \to \mathbb{R}$ defined by $f(x) = d(x, x_0)$ for all x in X is continuous. Is f uniformly continuous?

5. Show that the definition of uniform continuity can be rephrased as follows:

   To every real number $\epsilon > 0$, there exists real number $\delta > 0$ such that for all sets $E \subseteq X$ with diam $(E) < \delta$, we have diam $(f(E)) < \epsilon$.

6. Prove that the image of a Cauchy sequence under a uniformly continuous mapping is a cauchy sequence.

7. Prove that the image of a bounded set under a uniformly continuous mapping is a bounded set.

8. Show that if (X, d) is a metric space, and $f : X \to X$ is a continuous mapping of X into itself, then the mapping

$$\rho(x, y) = d(x, y) + \sum_{i=1}^{\infty} \frac{d(f^i(x), f^i(y))}{2^i}$$

where $f^i(x) = f(f^{i-1}(x))$ for $i \geq 1$, and $f^0(x) = x$, defines a metric on X which is equivalent to d, and that the mapping f is uniformly continuous with respect to ρ. Check that if (X, d) is totally bounded, then (X, ρ) is totally bounded.

9. Let (X, d_1) and (Y, d_2) be metric spaces. A mapping $f : X \to Y$ is uniformly continuous if and only if $d_2(f(A), f(B)) = 0$ whenever $d(A, B) = 0$ for all subsets A and B of X.

4.9 Semi-continuous Mappings

The real valued mappings that arise in applications may not be continuous. At a point these mappings satisfy a weaker condition than that required by the notion of continuity. If (X, d) is a metric space, and $x_0 \in X$, then the notion of continuity of a mapping $f : X \to \mathbb{R}$ at x_0 can be split into two parts:

(i) $d(x, x_0) < \delta$ implies that $f(x_0) - \epsilon < f(x)$, and

(ii) $d(x, x_0) < \delta$ implies that $f(x) < f(x_0) + \epsilon$.

If a mapping $f : X \to \mathbb{R}$ satisfies the condition (1) above, then f is said to be lower semi-continuous at x_0, and if f satisfies the condition (2), then f is said to be upper semi-continuous at x_0. Clearly, a mapping f is continuous if and only if it is both lower semi-continuous and upper semi-continuous. However, for technical reasons, we use extended real number system $\widetilde{\mathbb{R}}$ in place of $\mathbb{R}$, and the formal definitions of lower semi-continuous and upper semi-continuous mappings follow.

Definition 4.9.1 Let (X, d) be a metric space. An extended real valued mapping f defined on X is said to be upper semi-continuous at a point $x_0 \in X$ if for each extended real number u such that $f(x_0) < u$, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) < u.$$

The mapping f is called upper semi-continuous on X if it is upper semi-continuous at every point of X. Dually, f is said to be lower semi-continuous at x_0 if for each extended real number v such that $f(x_0) > v$, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that } f(x) > v.$$

The mapping f is called lower semi-continuous on X if it is lower semi-continuous at every point of X.

Since no metric has been defined on $\widetilde{\mathbb{R}}$, the above definition of upper semi-continuous and lower semi-continuous mappings can not be said to be a generalization of the notion of continuity. However, it can be shown (see references [18,33,36]) that a metric d exists on $\widetilde{\mathbb{R}}$ such that the collection $\mathbb{B}_l \cup \mathbb{B}_r \cup \theta$, where

$$\mathbb{B}_l = \{[-\infty, a[: a \in \mathbb{R} \quad \text{and } [-\infty, a[= \{x \in \widetilde{\mathbb{R}} : x < a\} \text{ known as left open ray}\},$$

$$\mathbb{B}_r = \{]a, \infty] : a \in \mathbb{R} \quad \text{and }]a, \infty] = \{x \in \widetilde{\mathbb{R}} : x > a\} \text{ known as right open ray}\},$$

$$\theta = \{]a, b[: a, b \in \mathbb{R} \quad \text{and } a < b\}$$

is an open base for the metric space $(\widetilde{\mathbb{R}}, d)$. This metric space is known as the extended real line. The real line $\mathbb{R}$ is clearly a subspace of $(\widetilde{\mathbb{R}}, d)$. But no "natural" definition of d is available. With this metric d on $\widetilde{\mathbb{R}}$ in mind, the discussion in this section progresses.

Theorem 4.9.2 Let (X, d) be a metric space, and $f : X \to \widetilde{\mathbb{R}}$ be a mapping. Let x_0 be a point in X such that $-\infty < f(x_0) < \infty$. Then f is upper semi-continuous at x_0 if and only if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) < f(x_0) + \epsilon.$$

Proof. Let $u = f(x_0)+\epsilon$. Then $f(x_0) < u$. Thus, by the definition of upper semi-continuous mapping at x_0, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) < u = f(x_0) + \epsilon.$$

Conversely, let u be an extended real number such that $f(x_0) < u$. Then there exists a real number $\epsilon > 0$ such that $f(x_0) + \epsilon < u$. By our assumption, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) < f(x_0) + \epsilon.$$

It follows from the definition of upper semi-continuity of a mapping at a point x_0 that if a mapping takes the value $+\infty$ at x_0, then it is upper semi-continuous at x_0.

Theorem 4.9.3 Let (X, d) be a metric space, and $f : X \to \widetilde{\mathbb{R}}$ be a mapping. Let x_0 be a point in X such that $-\infty < f(x_0) < \infty$. Then f is lower semi-continuous at x_0 if and only if for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) > f(x_0) - \epsilon.$$

Proof. The proof is similar to the one in the preceding theorem.

It may be noted that if a mapping $f : X \to \widetilde{\mathbb{R}}$ takes the values $-\infty$ at a point $x_0 \in X$, then, by the definition of lower semi-continuous mapping, f is lower semi-continuous at x_0.

Theorem 4.9.4 Let (X, d) be a metric space, and $f : X \to \widetilde{\mathbb{R}}$ be a mapping such that $-\infty < f(x_0) < \infty$ at a point $x_0 \in X$. Then f is continuous at x_0 if and only if f is both upper semi-continuous and lower semi-continuous at x_0 .

Proof. Let f be continuous at x_0, and $\epsilon > 0$ be a real number. Then the left open ray $]f(x_0) - \epsilon, \infty]$ is a neighbourhood of $f(x_0)$. By Theorem 4.1.6, there exists a neighbourhood M of x_0 such that $f(M) \subseteq]f(x_0) - \epsilon, \infty]$. Therefore, there exists a real number $\delta > 0$ such that the open sphere $S(x_0, \delta)$ is contained in M, and so $f(S(x_0, \delta)) \subseteq]f(x_0) - \epsilon, \infty]$. Thus,

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) > f(x_0) - \epsilon.$$

This means that f is lower semi-continuous at x_0. Similarly, f is upper semi-continuous at x_0. Conversely, we show that f is continuous at x_0. Let N be an open neighbourhood of $f(x_0)$. Then there exists an open interval $]a, b[$ such that $f(x_0) \in]a, b[\subseteq N$. Let $\epsilon_1 = f(x_0) - a$. Since f is lower semi-continuous at x_0, there exists a real number $\delta_1 > 0$ such that

$$d(x, x_0) < \delta_1 \quad \text{implies that} \quad f(x) > f(x_0) - \epsilon_1 = a.$$

Similarly, since f is upper semi-continuous at x_0, there exists a real number $\delta_2 > 0$ such that

$$d(x, x_0) < \delta_2 \quad \text{implies that} \quad f(x) < b.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) \in]a, b[\subseteq N.$$

Thus, there exists an open neighbourhood $S(x_0, \delta)$ of x_0 such that $f(S(x_0, \delta)) \subseteq N$. By Theorem 4.1.6, f is continuous at x_0.

Example 4.9.5 Let $f : \mathbb{R} \to \widetilde{\mathbb{R}}$ be a function defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational,} \\ -1 & \text{if } x \text{ is irrational.} \end{cases}$$

Then it is easily seen that $f(x)$ is upper semi-continuous at every rational number, and lower semi-continuous at every irrational number.

Example 4.9.6 Let $f : \mathbb{R} \to \widetilde{\mathbb{R}}$ be a function defined by

$$f(x) = \begin{cases} 0 & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$$

Clearly, f is upper semi-continuous at $x = 0$, f is not lower semi-continuous at $x = 0$ and, of course, f is not continuous at $x = 0$. At rest of the points, since f is continuous, it is both lower semi-continuous and upper semi-continuous.

Theorem 4.9.7 Let (X, d) be a metric space. A mapping $f : X \to \widetilde{\mathbb{R}}$ is upper semi-continuous at a point $x_0 \in X$ if and only if $\lim_{n\to\infty} \sup f(x_n) \leq f(x_0)$ for each sequence $\langle x_n \rangle$ in X converging to x_0. Dually, f is lower semi-continuous at x_0 if and only if $\lim_{n\to\infty} \inf f(x_n) \geq f(x_0)$ for each sequence $\langle x_n \rangle$ in X converging to x_0.

Proof. We prove the first assertion, the proof of the dual result is left to the reader. If x_0 is such that $f(x_0) = +\infty$, then the assertion holds trivially. So we assume that $f(x_0) < +\infty$. Suppose that f is upper semi-continuous at x_0. Let$\langle x_n \rangle$ be a sequence in X converging to x_0. Let $\epsilon > 0$ be a real number. Then, by Theorem 4.9.2, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) < f(x_0) + \epsilon.$$

Since $\langle x_n \rangle$ converges to x_0 in X, there exists an $m \in \mathbb{Z}_+$ such that $d(x_n, x_0) < \delta$ for all $n \geq m$. This means that $f(x_n) < f(x_0) + \epsilon$ for all $n \geq m$. Hence, $\lim_{x\to\infty} \sup f(x_n) \leq f(x_0) + \epsilon$. Since ϵ is arbitrary, $\lim_{n\to\infty} \sup f(x_n) \leq f(x_0)$. Conversely, assume that f is not upper semi-continuous at x_0. Then there exists a real number $\epsilon_0 > 0$ such that for each real number $\delta > 0$, there is a point x in X such that

$$d(x, x_0) < \delta \quad \text{but} \quad f(x) \geq f(x_0) + \epsilon_0.$$

In particular, for each $n \in \mathbb{Z}_+$, there exists a point, say x_n in X such that

$$d(x_n, x_0) < \frac{1}{n} \quad \text{but} \quad f(x_n) \geq f(x_0) + \epsilon_0.$$

Then $\langle x_n \rangle$ is a sequence in X converging to x_0 in X, but the sequence $\langle f(x_n) \rangle$ is bounded below by $f(x_0) + \epsilon_0$. Hence, it follows that $\lim_{n\to\infty} \sup f(x_n) \geq f(x_0) + \epsilon > f(x_0)$, a contradiction.

Theorem 4.9.8 Let (X, d) be a metric space. A mapping $f : X \to \widetilde{\mathbb{R}}$ is lower semi-continuous on X if and only if for each real number s, the set $\{x \in X : f(x) > s\}$ is open in X. Dually, f is upper semi-continuous on X if and only if for each real number u, the set $\{x \in X : f(x) < u\}$ is open in X.

Proof. We prove the first assertion, and leave the second to the reader. Let f be lower semi-continuous on X. Let s be a real number, and $A = \{x \in X : f(x) > s\}$. If A is empty, then it is open. So let x_0 be a point in A. Then $f(x_0) > s$. Since f is lower semi-continuous at x_0, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) > s.$$

Therefore, the open sphere $S(x_0, \delta)$ is contained in A, and A is open in X. Conversely, let x_0 be a point in X, and v be an extended real number such that $f(x_0) > v$. If v is finite, then, by our assumption, the set $A = \{x \in X : f(x) > v\}$ is open in X. Since $x_0 \in A$, there exists an open sphere $S(x_0, \delta)$ contained in A. Thus, for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) > v.$$

If $v = -\infty$, then there exists a real number v' such that $f(x_0) > v'$. Hence, there exists a real number $\delta > 0$ such that for all x in X

$$d(x, x_0) < \delta \quad \text{implies that} \quad f(x) > v' > v.$$

Corollary 4.9.9 Let (X, d) be a metric space. If $\{f_\alpha\}_{\alpha\in\Lambda}$ be a family of upper semi-continuous mappings $f_\alpha : X \to \widetilde{\mathbb{R}}$, then the mapping $g : X \to \widetilde{\mathbb{R}}$ defined by

$$g(x) = \inf\{f_\alpha(x) : \alpha \in \Lambda\}$$

is upper semi-continuous. Dually, if $\{f_\alpha\}_{\alpha\in\Lambda}$ is a family of lower semi-continuous mappings $f_\alpha : X \to \widetilde{\mathbb{R}}$, then the mapping $h : X \to \widetilde{\mathbb{R}}$ defined by

$$h(x) = \sup\{f_\alpha(x) : \alpha \in \Lambda\}$$

is lower semi-continuous.

Proof. As usual, we prove the first assertion. If u is a real number, then $g(x) < u$ if and only if there exists $\alpha \in \Lambda$ such that $f_\alpha(x) < u$. This means that

$$\{x \in X : g(x) < u\} = \bigcup_{\alpha\in\Lambda} \{x \in X : f_\alpha(x) < u\}.$$

Since f_α is upper semi-continuous for each α, $\{x \in X : f_\alpha(x) < u\}$ is an open set. Therefore, $\{x \in X : g(x) < u\}$ is an open set. Thus, g is upper semi-continuous by Theorem 4.9.8.

Corollary 4.9.10 Let (X, d) be a metric space. Let $\langle f_n \rangle$ be a sequence of monotonically decreasing (increasing) real valued continuous mappings. Then the mapping f on X defined by $f(x) = \lim_{n\to\infty} f_n(x)$ is upper (lower) semi-continuous.

Proof. If $\langle f_n\rangle$ is monotonically decreasing, then $f(x) = \lim_{n\to\infty} f_n(x) = \inf\{f_n(x) : n \in \mathbb{Z}_+\}$. If $\langle f_n\rangle$ is monotonically increasing, then $f(x) = \lim_{n\to\infty} f_n(x) = \sup\{f_n(x) : n \in \mathbb{Z}_+\}$. Now, the assertion follows from the preceding corollary.

Theorem 4.9.11 (Baire) For each real valued lower semi-continuous mapping g defined on a metric space (X, d), there exists a monotonically increasing sequence $\langle f_n\rangle$ of real valued continuous mappings defined on X that converges pointwise to g, that is, $g(x) = \lim_{n\to\infty} f_n(x)$. Dually, for each real valued upper semi-continuous mapping h defined on a metric space (X, d), there exists a monotonically decreasing sequence $\langle f_n\rangle$ of real valued continuous mappings defined on X that converges pointwise to h, that is, $h(x) = \lim_{n\to\infty} f_n(x)$.

Proof. First assume that g is bounded and takes its values in the open interval $]0, 1[$. For each $n \in \mathbb{Z}_+$, and $i = 1, 2, \ldots, n$, let $G_{i,n} = \{x \in X : g(x) > \frac{i}{n}\}$. Since g is lower semi-continuous, $G_{i,n}$'s are open sets in X. These sets are obviously nested, that is $G_{1,n} \supseteq G_{2,n} \supseteq \cdots \supseteq G_{n,n}$, and $G_{n,n}$ is empty set. By Theorem 4.4.11, for each $i = 1, 2, \ldots, n$, there exists a sequence $\langle f_m^{i,n}\rangle_{m=1}^{\infty}$ of real valued monotonically increasing continuous mappings defined on X which converges pointwise to the characteristic mapping $k_{i,n} = \chi_{G_{i,n}}$. Let us define

$$k_n = \frac{1}{n}\{k_{1,n} + k_{2,n} + \cdots + k_{n-1,n}\}$$

and

$$g_m^n = \frac{1}{n}\{f_m^{1,n} + f_m^{2,n} + \cdots + f_m^{n,n}\}$$

for all $m, n \in \mathbb{Z}_+$. The mappings g_m^n are all continuous on X. Also the sequence $\langle g_m^n\rangle_{m=1}^{\infty}$ is monotonically increasing, and converges pointwise to k_n. Now if $x \in G_{i,n} - G_{i+1,n}$ for some $i = 1, 2 \ldots, n-1$, then $\frac{i}{n} < g(x) \leq \frac{(i+1)}{n}$ and $k_n(x) = \frac{i}{n}$. Again, if x is not in $G_{1,n}$, then $g(x) \leq \frac{1}{n}$, and $k_n(x) = 0$. Thus, for each point x in X

$$k_n(x) < g(x) \leq k_n(x) + \frac{1}{n}.$$

Therefore, the limit $\lim_{m\to\infty} g_m^n(x) = k_n(x) \geq g(x) - \frac{1}{n}$. Let

$$g_m(x) = \max\{g_m^1(x), g_m^2(x), \ldots, g_m^m(x)\} \quad \text{for all } x \text{ in } X$$

Then g_m is a real valued and continuous mapping on X. Also, $g_m \leq g$. To see this, let x be any point in X. For each $i = 1, 2 \ldots, m$

$$g_m^i(x) = \frac{(f_m^{1,i}(x) + f_m^{2,i}(x) + \cdots + f_m^{i,i}(x))}{i}.$$

Since for each j, $1 \leq j \leq i$, the sequence $f_m^{j,i}(x) >_{m=1}^{\infty}$ is monotonically increasing sequence converging to the point $k_{j,i}(x) = XG_{j,i}(x)$, $f_m^{j,i}(x) \leq 1$. If $x \in G_{j-1,i} - G_{j,i}$, then $g_m^i(x) \leq \frac{(j-1)}{i}$ while $g(x) > \frac{(j-1)}{i}$. If x is not in $G_{1,i}$, then $g_m^i(x) \leq 0$ since $f_m^{j,i}(x) \leq 0$ for each $j = 1, 2 \ldots, i$, and $g(x) \in]0, 1[$. Thus, in any case, $g_m^i(x) \leq g(x)$. The sequence $\langle g_m\rangle$ is clearly monotonically increasing, and its pointwise limit which is same as its pointwise supremum is atleast as great as that of the sequence $\langle g_m^n\rangle$ for any given index n (for $m > n$, $g_m = g_m^1 \vee g_m^2 \vee \cdots \vee g_m^n \vee \cdots \vee g_m^m$, that is, $g_m^n \leq g_m$). Therefore, $\lim_{m\to\infty} g_m(x) = g(x)$ for each x in X.

Suppose now that g is bounded. We can find positive real numbers a and b such that $\hat{g} = ag+b$, and $\hat{g}$ takes the values in the interval $]0,1[$. Now $\hat{g}$ is a lower semi-continuous mapping on X. Therefore, there exists a monotonically increasing sequence $\langle \hat{g}_m \rangle$ of real valued continuous mappings defined on X which converges pointwise to $\hat{g}$. If we set $g_m = \frac{1}{a}(\hat{g}_m - b)$, then $\langle g_m \rangle$ is a sequence of real valued monotonically increasing continuous mappings which converges pointwise to g.

Finally, suppose that g is an arbitrary real valued lower semi-continuous mapping defined on X. Let $\hat{\phi} : \mathbb{R} \to]-1,1[$ be the standard homeomorphism $\hat{\phi}(t) = \frac{t}{(1+t)}$. Then the composition

$$\hat{g} = \hat{\phi} \circ g = \frac{g}{(1+|g|)}$$

is lower semi-continuous on X and bounded. So there exists a sequence $\langle \hat{g}_m \rangle$ of real valued monotone increasing continuous mappings which converges pointwise to $\hat{g}$. Also we can assume that each $\hat{g}_m$ is bounded below by -1 (since we may replace $\hat{g}_m$ by $\hat{g}_m \vee -1$). Since $\hat{g}_m \leq \hat{g}$, $\hat{g}_m(x) < 1$ for all x in X. However, $\hat{g}_m$ can still take value -1 at various points of X. In order to avoid this, we define

$$\hat{g}'_m = \frac{1}{2}\hat{g}_m + \frac{1}{4}\hat{g}_{m+1} + \cdots = \sum_{n=1}^{\infty} \frac{1}{2^n}\hat{g}_{m+n-1}.$$

Since the series on the right of the above equality is uniformly convergent series, $\hat{g}'_m$ is continuous mapping on X for each m in $\mathbb{Z}_+$. Moreover, it is easy to see that

$$\hat{g}_m \leq \hat{g}'_m \leq \hat{g}'_{m+1} \leq \hat{g}$$

for each $m \in \mathbb{Z}_+$ since $\hat{g}'_m \geq \hat{g}_m \sum_{n=1}^{\infty} \frac{1}{2^n}$, $\hat{g}'_m \leq \hat{g} \sum_{n=1}^{\infty} \frac{1}{2^n}$, and $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$.

Therefore, the sequence $\langle \hat{g}'_m \rangle$ is a monotonically increasing, and converges pointwise to $\hat{g}$. Also $\hat{g}'_m(x) = \hat{g}_m(x)$ when and only when $\hat{g}_m(x) = \hat{g}_{m+1}(x) = \ldots$, and this means that $\hat{g}_m(x) = \hat{g}(x) > -1$. Thus, each $\hat{g}'_m$ takes its values in the open interval $]-1,1[$, and we define the composition

$$g_m = (\hat{g})^{-1} \circ \hat{g}'_m = \frac{\hat{g}'_m}{(1-|\hat{g}'_m|)}$$

for all $m \in \mathbb{Z}_+$. Then $\langle g_m \rangle$ is the sequence of real valued monotonically increasing continuous mappings which converges pointwise to g.

Theorem 4.9.12 (Hahn Interpolation) Let f and g be real valued upper semi-continuous and lower semi-continuous mappings defined on a metric space X, respectively, such that $f(x) \leq g(x)$ for each x in X. Then there exists a continuous real valued mapping h defined on X such that $f \leq h \leq g$.

Proof. By Theorem 4.9.11 there exists a sequence $\langle f_n \rangle$ of real valued monotonically decreasing continuous mappings defined on X which converges pointwise to f, and there exists a sequence $\langle g_n \rangle$ of real valued monotonically increasing continuous mappings defined on X which converges pointwise to g. We define a third sequence $\langle p_n \rangle$ as follows:

$$p_{2n-1} = f_n - g_n, \quad p_{2n} = f_n - g_{n+1} \quad \text{for all } n \in \mathbb{Z}_+.$$

Then $\langle p_n \rangle$ is easily seen to be a monotonically decreasing sequence of continuous mappings which converges pointwise to the difference $f - g$. Now consider the series

$g_1 + \sum_{n=1}^{\infty}(-1)^{n+1}p_n$. Let $\langle s_n \rangle$ be its sequence of partial sums. Then $s_{2n-1} = g_n$ and $s_{2n} = f_n$, so that the sequence $\langle s_n \rangle$ is $\langle g_1, f_1, g_2, f_2, \ldots \rangle$. Now let

$$q_n = p_n \vee 0 \quad \text{for all } n \in \mathbb{Z}_+$$

and consider the series

$$g_1 + \sum_{n=1}^{\infty}(-1)^{n+1}q_n. \tag{A}$$

As the sequence $\langle q_n \rangle$ is monotonically decreasing and converges pointwise to $(f-g)^+ = (f-g) \vee 0 = 0$ (since $f(x) \leq g(x)$ for all x in X), the alternating series test shows that the series $g_1 + \sum_{n=1}^{\infty}(-1)^{n+1}q_n$ is pointwise convergent. For each x in X, let us denote the pointwise limit of $g_1 + \sum_{n=1}^{\infty}(-1)^{n+1}q_n$ by $h(x)$, that is,

$$h(x) = g_1(x) + \sum_{n=1}^{\infty}(-1)^{n+1}q_n(x) \quad \text{for all } x \text{ in } X.$$

Since $\langle p_n \rangle$ is monotonically decreasing, it is possible that $\langle p_n(x) \rangle$ is eventually negative and that $n = 2k-1$ is the first index such that $p_n(x) < 0$. Then $q_n(x) = 0$ for all $n \geq 2k-1$. Also $s_{2k-1}(x) = g_k(x)$, $f_k(x) \leq g_k(x)$, and

$$h(x) = g_1(x) + \sum_{n=1}^{2k-2}(-1)^{n+1}q_n(x) = g_1(x) + \sum_{n=1}^{2k-2}(-1)^{n+1}p_n(x) = s_{2k-1}(x).$$

Thus,

$$f(x) \leq f_k(x) \leq g_k(x) = h(x) \leq g(x).$$

If the first negative term in the sequence $\langle p_n(x) \rangle$ occurs for $n = 2k$, then $h(x) = g_1(x) + \sum_{n=1}^{2k-1}(-1)^{n+1}q_n(x) = s_{2k}(x)$, $f_k(x) \leq g_{k+1}(x)$, and $s_{2k}(x) = f_k(x)$. Thus,

$$f(x) \leq f_k(x) = h(x) \leq g_{k+1}(x) \leq g(x).$$

Finally, if $\langle p_n(x) \rangle$ is non-negative for each $n \in \mathbb{Z}_+$, $q_n(x) = p_n(x)$ for all $n \in \mathbb{Z}_+$, and we have

$$h(x) = f(x) = g(x).$$

Therefore, $f(x) \leq h(x) \leq g(x)$ for all x in X. The striking feature of the entire construction is that h is the pointwise limit of the alternating series in (A), it is also the pointwise limit of the increasing sequence of odd partial sums, and the pointwise limit of the decreasing sequence of even partial sums of (A). Thus, h is a real valued and both lower semi-continuous and upper semi-continuous by Corollary 4.9.9, and is therefore continuous by Theorem 4.9.4.

As the notion of continuity of a real valued mapping can equivalently be expressed in terms of two dual notions of lower semi-continuity and upper semi-continuity, we can split the notion of limit of a real valued mapping into two dual notions. Here again, it is convenient to use extended real valued mappings.

Definition 4.9.13 Let (X, d) be a metric space, and A be a subset of X. Let $f : A \to \widetilde{\mathbb{R}}$ be a mapping. Let x_0 be a point in $\bar{A}$. The limit superior of $f(x)$ as x in A tends to x_0, denoted by $\limsup_{\substack{x \to x_0 \\ x \in A}} f(x)$, is defined as

$$\limsup_{\substack{x \to x_0 \\ x \in A}} f(x) = \inf_{\epsilon > 0}\{M(f, x_0, \epsilon)\}$$

where $M(f, x_0, \epsilon) = \sup\{f(x) : x \in S(x_0, \epsilon) \cap A\}$, and ϵ is a positive real number.

Dually, the limit inferior of $f(x)$ as x in A tends to x_0, denoted by $\liminf_{\substack{x \to x_0 \\ x \in A}} f(x)$, is defined as

$$\liminf_{\substack{x \to x_0 \\ x \in A}} f(x) = \sup_{\epsilon > 0}\{m(f, x_0, \epsilon)\}$$

where $m(f, x_0, \epsilon) = \inf\{f(x) : x \in S(x_0, \epsilon) \cap A\}$. When $A = X$, we simply write $\limsup_{x \to x_0} f(x)$ and $\liminf_{x \to x_0} f(x)$ in the above two cases, respectively.

Since $m(f, x_0, \epsilon) \leq M(f, x_0, \epsilon)$, it follows that

$$\liminf_{\substack{x \to x_0 \\ x \in A}} f(x) \leq \limsup_{\substack{x \to x_0 \\ x \in A}} f(x)$$

at every point $x_0 \in \bar{A}$. If x_0 is a point in A, then

$$\liminf_{\substack{x \to x_0 \\ x \in A}} f(x) \leq f(x_0) \leq \limsup_{\substack{x \to x_0 \\ x \in A}} f(x).$$

Also, if f is a real valued mapping on A, and x_0 is in $\bar{A}$, then $\lim_{\substack{x \to x_0 \\ x \in A}} f(x)$ exists if and only if $\liminf_{\substack{x \to x_0 \\ x \in A}} f(x)$ and $\limsup_{\substack{x \to x_0 \\ x \in A}} f(x)$ are both real numbers that are equal, and in which case

$$\lim_{\substack{x \to x_0 \\ x \in A}} f(x) = \liminf_{\substack{x \to x_0 \\ x \in A}} f(x) = \limsup_{\substack{x \to x_0 \\ x \in A}} f(x).$$

Theorem 4.9.14 Let (X, d) be a metric space, and A be a subset of X. Let $f : A \to \widetilde{\mathbb{R}}$ be a mapping. Let $x_0 \in \bar{A}$, and

$$m(x_0) = \liminf_{\substack{x \to x_0 \\ x \in A}} f(x), \qquad M(x_0) = \limsup_{\substack{x \in x_0 \\ x \in A}} f(x).$$

Then the following properties hold.

(1) $m(x) \leq f(x) \leq M(x)$ for all x in A.

(2) $m(x) \leq M(x)$ for all x in $\bar{A}$.

(3) M is upper semi-continuous mapping on $\bar{A}$, and it is such that f is upper semi-continuous at a point x_0 in A relative to A if and only if $f(x_0) = M(x_0)$. Also if h is an arbitrary upper semi-continuous mapping on $\bar{A}$ such that $f \leq h$ on A, then $M \leq h$ as well.

Dually, m is a lower semi-continuous mapping on $\bar{A}$, and it is such that f is lower semi-continuous at $x_0 \in A$ relative to A if and only if $f(x_0) = m(x_0)$. Also, if g is arbitrary lower semi-continuous mapping on $\bar{A}$ such that $g \leq f$ on A, then $g \leq m$ as well.

Proof. (3) We prove the dual part leaving the first part to the reader. We show that m is a lower semi-continuous mapping on $\bar{A}$. Let u be a real number. Let $B = \{x \in \bar{A} : m(x) > u\}$. Let a_0 be a point in B. Then $m(a_0) > u$. Then there exists a real number $\delta > 0$ such that

$$m(f, a_0, \delta) = \inf\{f(x) : x \in S(a_0, \delta) \cap A\} > u.$$

If a is any point in $S(a_0, \delta) \cap \bar{A}$, then there exists a real number $\delta' > 0$ such that $S(a, \delta') \subseteq S(a_0, \delta)$. Therefore,

$$\begin{aligned} m(a) \geq m(f, a, \delta') &= \inf\{f(x) : x \in S(a, \delta') \cap A\} \\ &\geq \inf\{f(x) : x \in S(a_0, \delta) \cap A\} = m(f, a_0, \delta) > u. \end{aligned}$$

Thus, $a \in B$, and $S(a_0, \delta) \cap \bar{A} \subseteq B$. Thus, B is an open set in $\bar{A}$, and m is lower semi-continuous mapping on $\bar{A}$ by Theorem 4.9.8.

Now let f be lower semi-continuous at $x_0 \in A$ relative to A. If $f(x_0) = -\infty$, then $m(x_0) = f(x_0)$. So let us assume that $f(x_0) > -\infty$. Then there exists a real number v such that $f(x_0) > v$. Since f is lower semi-continuous at x_0, there exists a real number $\delta > 0$ such that for all $x \in S(x_0, \delta) \cap A$, $f(x) > v$. Therefore, $m(x_0) \geq v$. Since v is arbitrary real number less that $f(x_0)$, $m(x_0) \geq f(x_0)$, and $m(x_0) = f(x_0)$ follows. Conversely, let $f(x_0) = m(x_0)$, where x_0 is a point of A. Let u be a real number such that $f(x_0) > u$. Then the set B as defined above is an open set in $\bar{A}$, and it contains x_0. But then $B \cap A$ is open in A, and $f(x) > u$ for all x in $B \cap A$ since $f(x) \geq m(x)$ on A. Therefore, f is lower semi-continuous at x_0. If $f(x_0) = -\infty$, then f is lower semi-continuous at x_0.

Finally, let g be any lower semi-continuous mapping defined on $\bar{A}$ such that $g \leq f$ on A. Let $a_0 \in \bar{A}$. Then

$$\liminf_{\substack{x \to a_0 \\ x \in A}} f(x) \geq \liminf_{\substack{x \to a_0 \\ x \in A}} g(x) \geq \liminf_{\substack{x \to a_0 \\ x \in A}} g(x).$$

Since g is lower semi-continuous on $\bar{A}$ relative to $\bar{A}$, $m(a_0) \geq g(a_0)$.

EXERCISE

1. Let A be a subset of a metric space. At which points the characteristic mapping χ_A is upper semi-continuous? Where is χ_A lower semi-continuous? Where is χ_A continuous?

2. Let (X, d) be a metric space, and $f : X \to \widetilde{\mathbb{R}}$ be an extended real valued mapping. Prove the following statements:

   (a) f is upper (lower) semi-continuous at $x_0 \in X$ if and only if $-f$ is lower (upper) semi-continuous at x_0. (we have $-f(x) = \mp\infty$ if $f(x) = \pm\infty$.)

   (b) Let $g : X \to \widetilde{\mathbb{R}}$ be an extended real valued mapping, and a and b be non-negative real numbers. Then $af + bg$ is upper (lower) semi-continuous at x_0 if and only if f and g both are upper (lower) semi-continuous at x_0.

   (c) $f \vee g$ and $f \wedge g$ are upper (lower) semi-continuous at $x_0 \in X$ if and only if f and g both are upper (lower) semi-continuous at x_0.

Chapter 5

Completeness

5.1 Cauchy Sequences and Complete Metric Spaces

In real analysis, the reader must have come across the concept of a Cauchy sequence where it is proved that a sequence is a Cauchy sequence if and only if it is convergent. Such an equivalence is no longer true when these notions are extended to metric spaces. Neverthless, the notion of Cauchy sequence is important in metric spaces. Here it serves as a basis for classifying metric spaces.

Definition 5.1.1 A sequence $\langle x_n \rangle$ in a metric space (X, d) is called a Cauchy sequence or satisfies the Cauchy criterion if for each real number $\epsilon > 0$, there exists a $p \in \mathbb{Z}_+$ such that

$$d(x_n, x_m) < \epsilon \quad \text{for all } n, m \geq p.$$

Equivalently, let $\langle x_n \rangle$ be a sequence and let $T_n = \{x_k : k \geq n\}$, $n \in \mathbb{Z}_+$. Then $\langle x_n \rangle$ is a Cauchy sequence if and only if $\lim_{n\to\infty} \text{diam}(T_n) = 0$, where $\text{diam}(T_n)$ is the diameter of T_n, in the real line $\mathbb{R}$.

The following theorem covers the basic facts about Cauchy sequences.

Theorem 5.1.2 Let (X, d) be a metric space. Then the following statements hold.

1. Every convergent sequence in X is a Cauchy sequence.
2. Every Cauchy sequence is bounded.
3. Every Cauchy sequence that possesses a convergent subsequence is itself convergent to the limit of that subsequence.

Proof. (1) Let $\langle x_n \rangle$ be a convergent sequence in X that converges to $x \in X$. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that $d(x_n, x) < \frac{\epsilon}{2}$ for all $n \geq p$. Then

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} \quad \text{for all } n, m \geq p.$$

This proves that $\langle x_n \rangle$ is a Cauchy sequence in X.

(2) Let $\langle x_n \rangle$ be a Cauchy sequence in X. Then there exists an $n_0 \in \mathbb{Z}_+$ such that

$$d(x_n, x_m) < 1 \quad \text{for all } n, m \geq n_0.$$

If we let $k = \max_{n<n_0}\{d(x_n, x_{n_0})\}$ then the entire sequence $\langle x_n\rangle$ is contained in the open sphere $S(x_{n_0}, k+1)$ which is a bounded set. Therefore, $\langle x_n\rangle$ is a bounded sequence.

(3) Suppose that $\langle x_n\rangle$ is a Cauchy sequence in X, and $\langle x_{n_k}\rangle_{k=1}^{\infty}$ be a convergent subsequence of $\langle x_n\rangle$ which converges to $x \in X$. Let $\epsilon > 0$ be a real number. Then there exist $n_0, k_0 \in \mathbb{Z}_+$ such that

$$d(x_n, x_m) < \frac{\epsilon}{2} \quad \text{for all } n, m \geq n_0$$

and

$$d(x_{n_k}, x) < \frac{\epsilon}{2} \quad \text{for all } k \geq k_0.$$

Let us take $p = \max\{n_0, k_0\}$. Then

$$d(x_n, x) \leq d(x_n, x_{n_p}) + d(x_{n_p}, x) < \epsilon \quad \text{for all } n \geq p$$

which proves that $\langle x_n\rangle$ is convergent, and it converges to x.

From the third part of the above theorem it follows that a Cauchy sequence can have at most one cluster point, and if it has a cluster point it must converge to that cluster point.

Definition 5.1.3 A metric space (X, d) is called complete if every Cauchy sequence in X converges to some point in X.

The celebrated Polish mathematician Stefan Banach was probably the first person to recognise the fundamental role played by the completeness of metric spaces in analysis. It is in recognition of his contribution that the complete normed spaces, that is, normed spaces which are complete as a metric space, are known as Banach spaces. The notion of completeness of a metric space is essentially a property of distance that entails Cauchy sequences to converge.

In the following example, we prove the completeness of the real line which is instrumental in proving the completeness of several other metric spaces.

Example 5.1.4 The real line $\mathbb{R}$ is complete. Let $\langle x_n\rangle$ be a Cauchy sequence in $\mathbb{R}$. Therefore, $\langle x_n\rangle$ is bounded. Then $\lim_{n\to\infty} \sup x_n$ and $\lim_{n\to\infty} \inf x_n$ both exist in $\mathbb{R}$. Let $x = \lim_{n\to\infty} \inf x_n$, and $y = \lim_{n\to\infty} \sup x_n$. Then $x \leq y$. Suppose, if possible, $x < y$. Let $s = y - x$. Since $\langle x_n\rangle$ is a Cauchy sequence, there exists an $n_0 \in \mathbb{Z}_+$ such that

$$|x_n - x_m| < \frac{s}{2} \quad \text{for all } n, m \geq n_0.$$

Since $x = \lim\limits_{n\to\infty} \inf x_n$,

$$x \leq x_n < x + \frac{s}{4} \quad \text{for infinitely many values of } n.$$

This means that there is a positive integer $u \geq n_0$ such that

$$x \leq x_u < x + \frac{s}{4}.$$

Again since $y = \lim\limits_{n\to\infty} \sup x_n$,

$$y - \frac{s}{4} < x_n \leq y \quad \text{for infinitely many values of } n.$$

So there is a positive integer $v \geq n_0$ such that

$$y - \frac{s}{4} < x_v \leq y.$$

Therefore, it follows that

$$|x_u - x_v| > \frac{s}{2}$$

whereas since $u, v \geq n_0$, we get $|x_u - x_v| < \frac{s}{2}$, a contradiction. Thus, $x = y$, and $\mathbb{R}$ is complete.

Aliter. If $\langle x_n \rangle$ is a Cauchy sequence in $\mathbb{R}$, it is bounded, and hence it has a convergent subsequence. But then $\langle x_n \rangle$ is convergent by Theorem 5.1.2.

Example 5.1.5 The complex plane $\mathbb{C}$ is complete. Let $\langle z_n \rangle$ be a Cauchy sequence in $\mathbb{C}$. Then the sequences $\langle \operatorname{Re} z_n \rangle$ and $\langle \operatorname{Im} z_n \rangle$ are Cauchy sequences in $\mathbb{R}$, and therefore, these are convergent in $\mathbb{R}$. Let $x = \lim_{n\to\infty} \operatorname{Re} z_n$ and $y = \lim_{n\to\infty} \operatorname{Im} z_n$. Then $\langle z_n \rangle$ is convergent in $\mathbb{C}$, and it converges to $z = x + iy$.

Example 5.1.6 The n-dimensional Euclidean space $\mathbb{R}^n$ is complete. Let $\langle a_k \rangle$, where a_k is an n-tuple in $\mathbb{R}^n$ given by $a_k = (x_{k1}, x_{k2}, \ldots, x_{kn})$ be a Cauchy sequence in $\mathbb{R}^n$. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that

$$d(a_l, a_m) < \epsilon \quad \text{for all } l, m \geq p.$$

That is,

$$\left[\sum_{i=1}^{n} (x_{li} - x_{mi})^2\right]^{\frac{1}{2}} < \epsilon \quad \text{for all } l, m \geq p.$$

For each positive integer $i = 1, 2, \ldots, n$, it means that

$$|x_{li} - x_{mi}| < \epsilon \quad \text{for all } l, m \geq p.$$

Thus, the sequence $\langle x_{ki} \rangle_{k=1}^{\infty}$ in the ith co-ordinates of $\langle a_k \rangle$ is a Cauchy sequence in the real line $\mathbb{R}$, and by the completeness of $\mathbb{R}$, $\langle x_{ki} \rangle_{k=1}^{\infty}$ converges in $\mathbb{R}$ to a point $x_i \in \mathbb{R}$. Assume that $a = (x_1, x_2, \ldots, x_n)$. We show that the sequence $\langle a_k \rangle$ converges to a in $\mathbb{R}^n$. Since the sequence $\langle x_{ki} \rangle_{k=1}^{\infty}$ converges to x_i, there exists a positive integer p_i such that

$$|x_{ki} - x_i| < \frac{\epsilon}{\sqrt{n}} \quad \text{for all } k \geq p_i.$$

Let $p_0 = \max\{p_1, p_2, \ldots, p_n\}$. Then it follows that

$$\left[\sum_{i=1}^{n} (x_{ki} - x_i)^2\right]^{\frac{1}{2}} < \epsilon \quad \text{for all } k \geq p_0.$$

Thus, $d(a_k, a) < \epsilon$ for all $k \geq p_0$, and $\mathbb{R}^n$ is complete.

Example 5.1.7 The n-dimensional unitary space $\mathbb{C}^n$ is complete. The proof is similar as in the above example.

Example 5.1.8 l_p-spaces are complete. Recall that for each real number $p \geq 1$, the l_p-space is the set of all real or complex sequences $\langle x_n \rangle$ such that $\sum_{n=1}^{\infty} |x_n|^p < +\infty$, and the distance is given by

$$d(x, y) = \left[\sum_{n=1}^{\infty} |x_n - y_n|^p\right]^{\frac{1}{p}} \quad \text{for all } x = \langle x_n \rangle \quad \text{and} \quad y = \langle y_n \rangle \quad \text{in } l_p.$$

Let $\langle t_n \rangle$, where $t_n = \langle x_i^n \rangle_{i=1}^{\infty}$ is in l_p, be a Cauchy sequence in l_p. Let $\epsilon > 0$ be a real number. Then there exists an $n_0 \in \mathbb{Z}_+$ such that

$$d(t_n, t_m) < \epsilon \quad \text{for all } n, m \geq n_0. \tag{A}$$

This shows that $|x_i^n - x_i^m| < \epsilon$ for all $m, n \geq n_0$, and consequently $\langle x_i^n \rangle_{n=1}^{\infty}$ is Cauchy in $\mathbb{R}$ (or $\mathbb{C}$). Since these spaces are complete, $\langle x_i^n \rangle_{n=1}^{\infty}$ converges to a point $x_i \in \mathbb{R}$ (or $\mathbb{C}$).

Also for each $k \in \mathbb{Z}_+$, the statement (A) gives

$$\sum_{i=1}^{k} |x_i^n - x_i^m|^p \leq \epsilon^p \quad \text{for all } n, m \geq n_0.$$

Letting m to tend to infinity in the above, it follows that

$$\sum_{i=1}^{k} |x_i^n - x_i|^p \leq \epsilon^p \quad \text{for all } n \geq n_0. \tag{B}$$

We need to prove that $t = \langle x_1, x_2, \ldots \rangle$ is in l_p. The inequalities (B) above and Minkowski's inequality show that

$$\begin{aligned}
\left[\sum_{i=1}^{k} |x_i|^p\right]^{\frac{1}{p}} &\leq \left[\sum_{i=1}^{k} (|x_i^{n_0} - x_i| + |x_i^{n_0}|)^p\right]^{\frac{1}{p}} \\
&\leq \left[\sum_{i=1}^{k} |x_i^{n_0} - x_i|^p\right]^{\frac{1}{p}} + \left[\sum_{i=1}^{k} |x_i^{n_0}|^p\right]^{\frac{1}{p}} \\
&\leq \epsilon + \left[\sum_{i=1}^{k} |x_i^{n_0}|^p\right]^{\frac{1}{p}}.
\end{aligned}$$

Since $t_{n_0} = \langle x_i^{n_0} \rangle_{i=1}^{\infty}$ is in l_p, the above inequality shows that $\left\langle \left[\sum_{i=1}^{k} |x_i|^p\right]^{\frac{1}{p}} \right\rangle_{k=1}^{\infty}$ is bounded, and it is monotonically increasing, and hence the series $\sum_{i=1}^{\infty} |x_i|^p$ is convergent. Thus, t is in l_p. Also, it is obvious from (B) that $\langle t_n \rangle$ converges to t, and the completeness of l_p-space is now established.

Example 5.1.9 The metric space (X, d) in Example 2.1.23 is complete. Recall that X is the set of all bounded real (or complex) sequences, and the metric d on X is defined by $d(\langle x_n \rangle, \langle x_n' \rangle) = \sup \{|x_n - x_n'| : n \in \mathbb{Z}_+\}$. Let $\langle y_n \rangle$, where

$$y_n = \langle x_{n1}, x_{n2}, \ldots, x_{nk}, \ldots \rangle \quad \text{for all } n \in \mathbb{Z}_+ \text{ is an element of } X$$

be a cauchy sequence in X. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that $d(y_n, y_m) < \epsilon$ for all $n, m \geq p$. This means that

$$\sup_{k \in \mathbb{Z}_+} |x_{nk} - x_{mk}| < \epsilon \quad \text{for all } n, m \geq p. \tag{A}$$

Therefore, for each $k \in \mathbb{Z}_+$ the sequence $\langle x_{nk} \rangle_{n=1}^{\infty}$ is a cauchy sequence in $\mathbb{R}$ (or $\mathbb{C}$), and using the completeness of $\mathbb{R}$ (or $\mathbb{C}$), the sequence $\langle x_{nk} \rangle_{n=1}^{\infty}$ converegs to a point $x_k \in \mathbb{R}$

(or $\mathbb{C}$). Thus, we get a sequence $y = \langle x_1, x_2, \ldots \rangle$. When we let m to tend to ∞ in statement (A) above, it follows that

$$|x_{nk} - x_k| \leq \epsilon \quad \text{for all } n \geq p. \tag{B}$$

Now by the inequalities

$$|x_k| \leq |x_k - x_{pk}| + |x_{pk}| \quad \text{for all } k \in \mathbb{Z}_+,$$

we have

$$\sup_{k \in \mathbb{Z}_+} \{|x_k|\} \leq \sup_{k \in \mathbb{Z}_+} \{|x_{pk} - x_k|\} + \sup_{k \in \mathbb{Z}_+} \{|x_{pk}|\}$$

$$\leq \epsilon + \sup_{k \in \mathbb{Z}_+} \{|x_{pk}|\}.$$

Since y_p is bounded, it follows that y is in X. Taking supremum over k in statement (B) shows that $\langle y_n \rangle$ converges to y, and (X, d) is complete.

Example 5.1.10 The space $\mathcal{C}[a, b]$ of real valued continuous functions in Example 2.1.10 is complete. Let $\langle f_n \rangle$ be a Cauchy sequence in X. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that

$$d(f_n, f_m) < \frac{\epsilon}{3} \quad \text{for all } n, m \geq p$$

that is,

$$|f_n(x) - f_m(x)| < \frac{\epsilon}{3} \quad \text{for all } n, m \geq p, \text{ and } x \in [a, b]. \tag{A}$$

Therefore, $\langle f_n(x) \rangle$ is a Cauchy sequence in the real line $\mathbb{R}$ for each $x \in [a, b]$. Since $\mathbb{R}$ is complete, for each $x \in [a, b]$, $\langle f_n(x) \rangle$ converges to a point $y \in \mathbb{R}$; let us define a function $f : [a, b] \to \mathbb{R}$ such that $f(x) = y$. We show that f is continuous on $[a, b]$, and the sequence $\langle f_n \rangle$ converges to f. In statement (A) we keep n fixed and let m to tend to ∞, it becomes

$$|f_n(x) - f(x)| \leq \frac{\epsilon}{3} \quad \text{for all } n \geq p, \text{ and } x \in [a, b]. \tag{B}$$

Now let x_0 be any point in $[a, b]$. Then

$$|f(x) - f(x_0)| \leq |f(x) - f_p(x)| + |f_p(x) - f_p(x_0)| + |f_p(x_0) - f(x_0)|$$

$$< \frac{\epsilon}{3} + |f_p(x) - f_p(x_0)| + \frac{\epsilon}{3}.$$

Since f_p is continuous at x_0, there exists a real number $\delta > 0$ such that

$$|f_p(x) - f_p(x_0)| < \frac{\epsilon}{3} \quad \text{whenever } |x - x_0| < \delta.$$

Therefore,

$$|x - x_0| < \delta \quad \text{implies that } |f(x) - f(x_0) < \epsilon.$$

Since x_0 is an arbitrary point, f is continuous on $[a, b]$. Also, by the statement (B), it follows that $\langle f_n \rangle$ converges to f.

Example 5.1.11 The metric space $\mathcal{C}([a,b],d)$ in Example 2.1.11 is not complete. To see this, we construct a Cauchy sequence $\langle f_n\rangle$ in $\mathcal{C}[a,b]$ which does not converge to any point in $\mathcal{C}[a,b]$. Let us define

$$f_n(t) = \begin{cases} 0 & \text{if } a \le t < \dfrac{(a+b)}{2} - \dfrac{(b-a)}{2n}, \\ \dfrac{2nt}{(b-a)} - \dfrac{n(a+b)}{(b-a)} + 1 & \text{if } \dfrac{(a+b)}{2} - \dfrac{(b-a)}{2n} \le t < \dfrac{(a+b)}{2}, \\ 1 & \text{if } \dfrac{(a+b)}{2} \le t \le b. \end{cases}$$

Each f_n is a continuous mapping defined on $[a,b]$. For $n > m$, the functions f_n and f_m are shown in Fig. 5.1 in which the point A is $(\frac{(a+b)}{2} - \frac{(b-a)}{2m}, 0)$ and B is $(\frac{(a+b)}{2} - \frac{(b-a)}{2n}, 0)$.

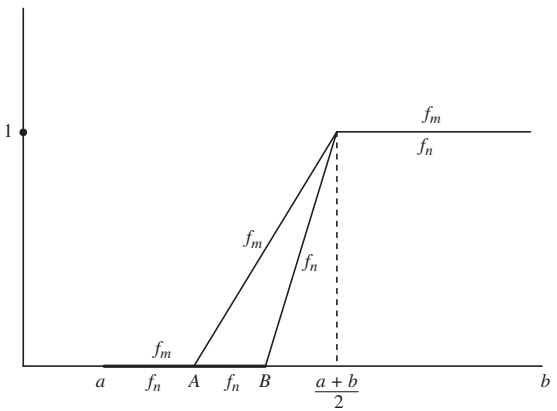


Fig. 5.1

For $n > m$, $f_n(t) \le f_m(t)$ for all $t \in [a,b]$, and

$$\begin{aligned} d(f_n, f_m) &= \int_a^b |f_n(t) - f_m(t)|\, dt = \int_a^b (f_m(t) - f_n(t))\, dt \\ &= \int_{\frac{(a+b)}{2} - \frac{(b-a)}{2m}}^{\frac{(a+b)}{2} - \frac{(b-a)}{2n}} f_m(t)\, dt + \int_{\frac{(a+b)}{2} - \frac{(b-a)}{2n}}^{\frac{(a+b)}{2}} (f_m(t) - f_n(t))\, dt \\ &< (b-a)\left(\frac{1}{m} + \frac{1}{n}\right). \end{aligned}$$

From the above inequality, it follows that $\langle f_n\rangle$ is a Cauchy sequence in $\mathcal{C}[a,b]$. Suppose that $\langle f_n\rangle$ converges to f in $\mathcal{C}[a,b]$. This means that

$$\lim_{n\to\infty} d(f_n, f) = \lim_{n\to\infty} \int_a^b |f_n(t) - f(t)| dt = 0.$$

That is,

$$\lim_{n\to\infty}\left[\int_a^{\frac{(a+b)}{2}-\frac{(b-a)}{2n}}|f(t)|dt+\int_{\frac{(a+b)}{2}-\frac{(b-a)}{2n}}^{\frac{(a+b)}{2}}|f_n(t)-f(t)|dt+\int_{\frac{(a+b)}{2}}^{b}|1-f(t)|dt\right]=0.$$

Since the integrands in the above integrals are non-negative, the limit of each of the integrals is zero, and we have

$$\lim_{n\to\infty}\int_a^{\frac{(a+b)}{2}-\frac{(b-a)}{2n}}|f(t)|dt=0 \quad \text{and} \quad \int_{\frac{(a+b)}{2}}^{b}|1-f(t)|dt=0.$$

This means that

$$f(t)=\begin{cases}0 & \text{if } a\le t<\frac{(a+b)}{2},\\ 1 & \text{if } \frac{(a+b)}{2}\le t\le b,\end{cases}$$

which is not a continuous function. This is a contradiction, and it proves our assertion.

Theorem 5.1.12 Let (X,d) be a metric space, and $\bar{d}$ be the standard bounded metric on X given by $\bar{d}(x,y)=\min\{d(x,y),1\}$. Then (X,d) is complete if and only if $(X,\bar{d})$ is complete.

Proof. First assume that (X,d) is complete. Let $\langle x_n\rangle$ be a Cauchy sequence in $(X,\bar{d})$. Let $\epsilon>0$ be a real number. Therefore, for $0<\epsilon'<\min\{\epsilon,1\}$, there exists a $p\in\mathbb{Z}_+$ such that $\bar{d}(x_n,x_m)<\epsilon'$ for all $n,m\ge p$. It means that

$$d(x_n,x_m)<\epsilon'<\epsilon \quad \text{for all } n,m\ge p. \tag{A}$$

Thus, $\langle x_n\rangle$ is a Cauchy sequence in (X,d). Since (X,d) is complete, $\langle x_n\rangle$ converges to a point x in X. Letting m tend to infinity in statement (A), it follows that

$$d(x_n,x)<\epsilon \quad \text{for all } n\ge p.$$

Now the completeness $d(X,\bar{d})$ follows from the inequalities

$$\bar{d}(x_n,x_m)\le d(x_n,x)<\epsilon \quad \text{for all } n\ge p.$$

Conversely, let $\langle x_n\rangle$ be a Cauchy sequence in (X,d). Let $\epsilon>0$ be a real number, and $0<\epsilon'<\min\{\epsilon,1\}$. Then there exists a $p\in\mathbb{Z}_+$ such that $d(x_n,x_m)<\epsilon'$ for all $n,m\ge p$. That is, $\bar{d}(x_n,x_m)<\epsilon'<\epsilon$ for all $n,m\ge p$. Thus, $\langle x_n\rangle$ is a cauchy sequence in $(X,\bar{d})$. By completeness of $(X,\bar{d})$, there is a point $x\in X$ such that $\langle x_n\rangle$ converges to x. Letting m tend to infinity

$$\bar{d}(x_n,x)\le\epsilon'<\epsilon \quad \text{for all } n\ge p.$$

That is, $d(x_n,x)\le\epsilon$ for all $n\ge p$, and the proof is complete.

Recall the metric space in Definition 2.15.9: Let J be any arbitrary index set, and $\{(x_\alpha,d_\alpha)\}_{\alpha\in J}$ be a family of metric spaces. Let $X=\prod_{\alpha\in J}X_\alpha$. Then the metric

$$\bar{\rho}(x,y)=\sup\{\bar{d}_\alpha(x_\alpha,y_\alpha):\alpha\in J\}$$

where $\bar{d}_\alpha(x_\alpha,y_\alpha)=\min\{1,d_\alpha(x_\alpha,y_\alpha)\}$, is called a uniform metric on X. With respect to this uniform metric, we have the following theorem.

Theorem 5.1.13 If $\{(X_\alpha, d_\alpha)\}_{\alpha\in J}$ is a family of complete metric spaces, then the metric space $(X, \bar{\rho})$ is complete.

Proof. Let $\langle f_n\rangle$ be a Cauchy sequence in X. Let $\epsilon > 0$ be a real number. Then for each $\alpha \in J$

$$\bar{d}_\alpha(f_n(\alpha), f_m(\alpha)) \leq \bar{\rho}(f_n, f_m).$$

This means that for each $\alpha \in J$, the sequence $\langle f_n(\alpha)\rangle$ is a Cauchy sequence in $(X_\alpha, \bar{d}_\alpha)$. Since $(X_\alpha, \bar{d}_\alpha)$ is complete by the preceding theorem, $\langle f_n(\alpha)\rangle$ converges to $f(\alpha)$ in $(X_\alpha, \bar{d}_\alpha)$. Thus, we get a mapping $f : J \to X$ such that $f(\alpha) \in X_\alpha$, and $f(\alpha)$ is the limit of the sequence $\langle f_n(\alpha)\rangle$ in X_α. Now there exists a $p \in \mathbb{Z}_+$ such that $\bar{\rho}(f_n, f_m) < \frac{\epsilon}{2}$ for all $n, m \geq p$, that is,

$$\bar{d}_\alpha(f_n(\alpha), f_m(\alpha)) < \frac{\epsilon}{2} \quad \text{for all } m, n \geq p \text{ and } \alpha \in J.$$

Now letting m to tend to infinity, we have

$$\bar{d}_\alpha(f_n(\alpha), f(\alpha)) \leq \frac{\epsilon}{2} \quad \text{for all } n \geq p \text{ and } \alpha \in J.$$

Therefore, $\bar{\rho}(f_n, f) < \epsilon$ for all $n \geq p$, and the proof is complete.

Example 5.1.14 Fréchet space F in Example 2.1.22 is complete. Let $\langle x_n\rangle$, where $x_n = \langle x_{n1}, x_{n2}, \ldots\rangle$ be a Cauchy sequence in F. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that $\sigma(x_n, x_m) < \epsilon$ for all $n, m \geq p$. Therefore,

$$\frac{|x_{ni} - x_{mi}|}{2^i(1 + |x_{ni} - x_{mi}|)} < \epsilon \quad \text{for all } m, n \geq p \text{ and for each } i.$$

This means that $|x_{ni} - x_{mi}| < 2^{1+i}\epsilon$ for all $m, n \geq p$. Thus, $\langle x_{ni}\rangle_{n=1}^{\infty}$ in the ith co-ordinates of $\langle x_n\rangle$ is a Cauchy sequence in $\mathbb{R}$ (or $\mathbb{C}$), and so converges to a point $x^i \in \mathbb{R}$ (or $\mathbb{C}$). If we let $x = \langle x^1, x^2, \ldots\rangle$, then x is in F. Since the co-ordinate-wise, convergence is equivalent to the pointwise convergence in Fréchet space by Theorem 3.3.3, the convergence of $\langle x_n\rangle$ to x follows, and thus F is complete.

Example 5.1.15 Let (X, d) be a metric space, and Z be a set. Recall Example 2.1.31, where the set $\mathbb{B}(Z, X)$ of all bounded mappings of Z into X is a metric space with metric ρ defined by

$$\rho(\phi, \psi) = \sup\{d(\phi(z), \psi(z)) : z \in Z\}.$$

We show that $\mathbb{B}(Z, X)$ is a complete metric space with metric ρ, if (X, d) is complete. Let $\langle\phi_n\rangle$ be a Cauchy sequence in $\mathbb{B}(Z, X)$. Let $\epsilon > 0$ be a real number. Then it follows that $\langle\phi_n(z)\rangle$ is a Cauchy sequence in X for each $z \in Z$, and consequently convergent in X. Let us denote the limit $\lim_{n\to\infty}\phi_n(z)$ by $\phi(z)$. Thus, we get a mapping $\phi : Z \to X$ given by $\phi(z) = \lim_{n\to\infty}\phi_n(z)$. We show that ϕ is in $\mathbb{B}(Z, X)$, and that the sequence $\langle\phi_n\rangle$ converges to ϕ. Since $\langle\phi_n\rangle$ is a Cauchy sequence, there exists an $n_0 \in \mathbb{Z}_+$ such that

$$\rho(\phi_n, \phi_m) < \frac{\epsilon}{2} \quad \text{for all } n, m \geq n_0.$$

This means that

$$d(\phi_n(z), \phi_m(z)) < \frac{\epsilon}{2} \quad \text{for all } n, m \geq n_0 \text{ and } z \in Z.$$

If m tends to infinity, then

$$d(\phi_n(z), \phi(z)) \leq \frac{\epsilon}{2} \quad \text{for all } n \geq n_0 \text{ and } z \in Z.$$

This means that $\rho(\phi_n, \phi) < \epsilon$ for all $n \geq m_0$. Thus, $\langle \phi_n \rangle$ converges to ϕ.

Also, the closed sphere $\{x \in X : d(\phi_{n_0}(z), x) \leq \epsilon\}$ is a closed set containing $\phi(z)$. This means that if the diameter $\operatorname{diam}(\phi_{n_0}(Z)) = M$, then $\operatorname{diam}(\phi(Z)) \leq M + 2\epsilon$, and ϕ is bounded.

Conversely, assume that (X, d) is not complete. Then there exists a Cauchy sequence $\langle x_n \rangle$ in X which does not converge to any point in X. We define a sequence $\langle \phi_n \rangle$ in $\mathbb{B}(Z, X)$ by setting $\phi_n(z) = x_n$ for all z in Z. It is easily seen that the sequence $\langle \phi_n \rangle$ is a Cauchy sequence in $\mathbb{B}(Z, X)$. We show that $\langle \phi_n \rangle$ does not converge in $\mathbb{B}(Z, X)$. Assume, on the contrary that $\langle \phi_n \rangle$ converges to a point ϕ in $\mathbb{B}(Z, X)$. Let $\epsilon > 0$ be a real number. Then there exists an $n_0 \in \mathbb{Z}_+$ such that $\rho(\phi_n, \phi) < \epsilon$ for all $n \geq n_0$. This means that $d(x_n, \phi(z)) < \epsilon$ for $n \geq n_0$ and $z \in Z$, and so $\langle x_n \rangle$ converges to $\phi(z)$ in X, a contradiction. Thus, if Z is non-empty, the metric space $\mathbb{B}(Z, X)$ with metric ρ is complete if and only if (X, d) is complete.

Example 5.1.16 Let (X, d) and (Y, ρ) be metric spaces. Then the set $\mathcal{C}_b(X, Y)$ of all continuous and bounded mappings from X into Y with metric

$$\sigma(f, g) = \sup_{x \in X} \{\rho(f(x), g(x))\}$$

is a complete metric space if and only if (Y, ρ) is complete. The argument is similar as in the preceding example.

EXERCISE

1. Let $\langle x_n \rangle$ and $\langle y_n \rangle$ be sequences in (X, d) such that $\langle x_n \rangle$ is a Cauchy sequence, and $d(x_n, y_n)$ tends to zero as n tends to infinity. Show that
   (a) $\langle y_n \rangle$ is a Cauchy sequence in (X, d).
   (b) $\langle y_n \rangle$ converges to a point y in X if and only if $\langle x_n \rangle$ converges to y.

2. Let $d : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ be defined by $d(m, n) = \frac{|m-n|}{mn}$. Show that d is a metric on $\mathbb{N}$. Show that d is equivalent to the usual metric on $\mathbb{N}$. Show that $(\mathbb{N}, d)$ is not complete.

3. Show that the open interval $]0, 1[$ with usual metric is not complete. Deduce that if (X, d_1) and (Y, d_2) are two homeomorphic metric spaces, then completeness of one metric space does not imply the completeness of the other metric space.

4. Let $\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$\rho(x, y) = \frac{|x - y|}{(\sqrt{1 + x^2})(\sqrt{1 + y^2})}.$$

   Show that $(\mathbb{R}, \rho)$ is a metric space which is not complete. Show also that the identity mapping $i : (\mathbb{R}, d) \to (\mathbb{R}, \rho)$, where d is the usual metric on $\mathbb{R}$, is uniformly continuous. (This shows that uniformly continuous mapping need not preserve completeness.)

5. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a uniformly continuous mapping such that if $\langle x_n \rangle$ is a Cauchy sequence in X, then $\langle f(x_n) \rangle$ is a Cauchy sequence in Y. Further, if f is bijective such that f^{-1} is also uniformly continuous, prove that (X, d_1) is complete if and only if (Y, d_2) is complete.

6. Let $\mathcal{C}[a, b]$ be the set of all continuous real valued functions defined on $[a, b]$. Show that the mapping ρ defined by

$$\rho(f, g) = \sup_{t \in [a,b]} \{e^{-k(t-a)} |f(t) - g(t)|\}$$

is a metric on $\mathcal{C}[a, b]$, and $(\mathcal{C}[a, b], \rho)$ is complete.

7. Let $\mathcal{C}[a, b]$ be the metric space in Example 2.1.11. Show that the sequence $\langle f_n \rangle$ in $\mathcal{C}[a, b]$ given by

$$f_n(t) = \begin{cases} 0 & \text{if } a \leq t \leq 0, \\ nt & \text{if } 0 < t < \frac{1}{n}, \\ 1 & \text{if } \frac{1}{n} \leq t \leq b \end{cases}$$

is a Cauchy sequence which does not converge in $\mathcal{C}[a, b]$. However, if $a = 0$, then show that the above example of the sequence $\langle f_n \rangle$ does not establish that $\mathcal{C}[a, b]$ is not complete.

8. Show that the two metrics on the set of real numbers $\mathbb{R}$ given by

$$d_1(x, y) = |x - y|,$$

$$d_2(x, y) = \frac{|x - y|}{1 + |x - y|^2}$$

are equivalent. Show that a sequence $\langle x_n \rangle$ is a Cauchy sequence in $(\mathbb{R}, d_1)$ if and only if $\langle x_n \rangle$ is a Cauchy sequence in $(\mathbb{R}, d_2)$. Also show that $\langle x_n \rangle$, where $x_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is a Cauchy sequence in $(\mathbb{R}, d_2)$ which does not converge. This shows that equivalent metrics need not preserve completeness.

9. Let (X, d) be a metric space. Define a metric d' on X by

$$d'(x, y) = \begin{cases} d(x, y) & \text{if } d(x, y) < 1, \\ 1 & \text{if } d(x, y) \geq 1. \end{cases}$$

Show that d' is a metric on X which is bounded. Also show that a sequence $\langle x_n \rangle$ in X is a Cauchy sequence in (X, d) if and only if $\langle x_n \rangle$ is a Cauchy sequences in (X, d'). Are the metrics d and d' equivalent?

10. Check that if (X, d) is a complete metric space, and a metric ρ on X is uniformly equivalent to d, then (X, ρ) is complete.

11. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a continuous and onto mapping such that $d_1(x_1, x_2) \leq \bar{d}_2(f(x_1), f(x_2))$ for all x_1, x_2 in X. If X is complete, then show that Y is complete.

12. Let (X, d) be a metric space, and $d(A, B) > 0$ for any pair of disjoint closed sets A and B in X. Prove that (X, d) is complete.

13. Show that the subset C of the convergent sequences in the metric space of bounded sequences in Example 2.1.23 is complete. Let C_0 be the subset of C consisting of sequences that converge to 0. Is C_0 complete?

14. Let $X = \{\langle x_n \rangle : x_n \in \mathbb{Z}_+ \text{ for all } n \in \mathbb{Z}_+\}$, and $d(\langle x_n \rangle, \langle y_n \rangle) = \frac{1}{n}$, where n is the smallest integer for which $x_n \neq y_n$. Prove that d is a metric on X, and that (X, d) is complete.

15. Give an example of a complete metric space (X, d), a non-complete metric space (Y, ρ) and an onto mapping $f : X \to Y$ such that f is uniformly continuous.

5.2 Completeness of Subspaces and Products

Theorem 5.2.1 A closed subspace of a complete metric space is complete. Conversely, a complete subspace of an arbitrary metric space (X, d) is a closed set in X.

Proof. Let (X, d) be a complete metric space, and F be a non-empty closed subset of X. Then any Cauchy sequence $\langle x_n \rangle$ in F is a Cauchy sequence in X, and thus converges to a point x in X by the completeness of X. Since F is a closed set, the limit of $\langle x_n \rangle$ is in F. This means that the sequence $\langle x_n \rangle$ converges to $x \in F$. Thus, F is complete. Conversely, let F be a complete subspace of a metric space (X, d). We show that F is a closed set in X. Let x be a point in X but not in F such that x is a limit point of F. Then we can extract a sequence $\langle x_n \rangle$ in F such that $\langle x_n \rangle$ converges to x in X. This means that $\langle x_n \rangle$ is a Cauchy sequence in F. Since F is complete, $\langle x_n \rangle$ converges to a point y in F. But then $\langle x_n \rangle$ converges to y in X. So $\langle x_n \rangle$ converges in X to two distinct points x and y, a contradiction. Therefore, F is a closed set in X.

Example 5.2.2 Let P be the set of all polynomial functions (of all degrees) defined on $[a, b]$. Let $d : P \to \mathbb{R}$ be a mapping defined by

$$d(f, g) = \max_{a \leq t \leq b} \{|f(t) - g(t)|\} \quad \text{for all } f \text{ and } g \text{ in } P.$$

Since every polynomial function is continuous, $P \subseteq \mathcal{C}[a, b]$. Also the metric d is the restriction to P of the sup metric on $\mathcal{C}[a, b]$ (see Example 2.1.10). Thus, (P, d) is a subspace of $\mathcal{C}[a, b]$. We show that (P, d) is not complete. Consider a sequence $\langle f_n \rangle$ in P given by

$$f_n(t) = 1 + \frac{(t-a)}{2} + \frac{(t-a)^2}{2^2} + \cdots + \frac{(t-a)^n}{2^n} \quad \text{for all } t \in [a, b] \text{ and } n \in \mathbb{Z}_+.$$

Now, for $n > m$

$$d(f_n, f_m) = \max_{a \le t \le b} \left\{ \left| \sum_{k=m+1}^{n} \frac{(t-a)^k}{2^k} \right| \right\}$$

$$\le \max_{a \le t \le b} \sum_{k=m+1}^{n} \left(\frac{|t-a|}{2} \right)^k \le (b-a) \sum_{k=m+1}^{n} \frac{1}{2^k}$$

$$= (b-a) \left(\frac{1}{2^m} - \frac{1}{2^n} \right)$$

and this is arbitrarily small for large enough values of m and n. This shows that $\langle f_n \rangle$ is a Cauchy sequence in P. However, the sequence $\langle f_n \rangle$ does not converge to any polynomial function in P since the limit

$$\lim_{n \to \infty} f_n(t) = \sum_{k=0}^{\infty} \frac{(t-a)^k}{2^k} = \frac{2}{2-t+a}$$

is not a polynomial function. Moreover, $f(t)$ is not even defined at $t = a+2$ if $a+2 \in [a, b]$.

Our next theorem states that the countable product of complete metric spaces is complete. Recall Theorem 2.15.10 and Theorem 5.1.12.

Theorem 5.2.3 Let $\{(X_i, d_i) : i \in J\}$ be a countable family of metric spaces. Then the product metric space (X, D), where

$$D(x, y) = \text{lub}_{i \in J} \{ \bar{d}_i(x_i, y_i)/i \}$$

for all $x = \langle x_i \rangle$ and $y = \langle y_i \rangle$ in $X = \prod_{i \in J} X_i$, and $\bar{d}_i(x_i, y_i) = \min\{d_i(x_i, y_i), 1\}$, is complete if and only if each factor space (X_i, d_i) is complete.

Proof. Let $y_n = (x_{nj})_{j \in J} \in X$ for each $n \in \mathbb{Z}_+$, and $\langle y_n \rangle$ be a Cauchy sequence in X. Let $\epsilon > 0$ be a real number. Therefore, there exists a $p \in \mathbb{Z}_+$ such that

$$D(y_n, y_m) < \frac{\epsilon}{2} \quad \text{for all } n, m \ge p.$$

That is,

$$\bar{d}_i(x_{ni}, x_{mi}) < \frac{\epsilon i}{2} \quad \text{for all } m, n \ge p \text{ and } i \in J. \tag{A}$$

Since ϵ is arbitrary, it follows that $\langle x_{ni} \rangle_{n=1}^{\infty}$ is a Cauchy sequence in $(X_i, \bar{d}_i)$ for each $i \in J$. Since $(X_i, \bar{d}_i)$ is complete, $\langle x_{ni} \rangle_{n=1}^{\infty}$ converges to a point $x_i \in X_i$. Let $y = (x_i)_{i \in J}$. Letting m to tend to infinity in statement (A), we get

$$\bar{d}_i(x_{ni}, x_i) \le \frac{\epsilon i}{2} \quad \text{for all } n \ge p \text{ and } i \in J.$$

This means that $D(y_n, y) < \epsilon$ for all $n \ge p$ so that (X, d) is complete.

Conversely, assume that the product metric space (X, D) is complete. Then the subspace $Y_j = \prod_{i \in J} A_i$, where $A_j = X_j$, and $A_i = \{x_i^*\} \subseteq X_i$ for $i \ne j$, is a closed set, and therefore it is complete. Now the projection mapping $p_j^* = p_j|_{Y_j} : Y_j \to X_j$ is a homeomorphism, and that for every Cauchy sequence $\langle x_n \rangle$ in $(X_j, \bar{d}_j)$, the sequence $\langle p_j^{*-1}(x_n) \rangle$ is a Cauchy sequence in Y_j. The image of the limit of $\langle p_j^{*-1}(x_n) \rangle$ is the limit of the sequence $\langle x_n \rangle$ so that $(X_j, \bar{d}_j)$ is complete.

5.3 Characterisations of Completeness

There are several conditions which are equivalent to completeness. The two of them can be immediately discussed.

Theorem 5.3.1 A metric space X is complete if and only if every Cauchy sequence in X has a convergent subsequence.

Proof. The proof follows from Theorem 5.1.2.

The following theorem is known as Cantor's intersection theorem.

Theorem 5.3.2 A metric space (X, d) is complete if and only if for any nested sequence $\langle F_n \rangle$ of non-empty closed sets in X with diameter of F_n tending to zero as n tends to infinity, the intersection $\bigcap_{i=1}^{\infty} F_n$ contains exactly one point.

Proof. Let(X, d) be a complete metric space, and $\langle F_n \rangle$ be a nested sequence of non-empty closed sets in X such that the diameter of F_n, $\operatorname{diam} F_n$, tends to zero as n tends to infinity. Let $F = \bigcap_{n=1}^{\infty} F_n$. Suppose that F contains at least two points x and y of X. Then x and y are in F_n for all positive integer n. Thus, $\operatorname{diam} F_n \geq d(x, y)$ for all positive integer n. But this contradicts our assumption that $\operatorname{diam} F_n$ tends to zero as n tends to infinity. Therefore, it remains to show that F is non-empty. Since each F_n is non-empty, we can choose a point x_n in F_n. We show that $\langle x_n \rangle$ is a Cauchy sequence in X. Let $\epsilon > 0$ be a real number. Since $\operatorname{diam} F_n$ tends to zero as n tends to infinity, there exists a $p \in \mathbb{Z}_+$ such that $\operatorname{diam} F_n < \epsilon$ for all $n \geq p$. Since $x_n \in F_n$, and $F_n \subseteq F_p$ for all $n \geq p$, $x_n \in F_p$ for all $n \geq p$. Therefore,

$$d(x_n, x_m) \leq \operatorname{diam} F_p < \epsilon \quad \text{for all } n, m \geq p.$$

This shows that $\langle x_n \rangle$ is a Cauchy sequence in X. Since X is complete, $\langle x_n \rangle$ converges to a point x in X. We show that x is in F. For this we show that x is in F_k for each positive integer k. Now the subsequence $\langle x_k, x_{k+1}, \ldots \rangle$ of $\langle x_n \rangle$ is in F_k, and it converges to x. Since F_k is a closed set, x is in F_k. Since k is arbitrary, $x \in \bigcap_{n=1}^{\infty} F_n$. Thus, F contains exactly one point.

Conversely, assume that $\langle x_n \rangle$ is a Cauchy sequence in X. Then there exists an n_1 such that

$$d(x_n, x_m) < \frac{1}{2} \quad \text{for all } n, m \geq n_1.$$

In particular,

$$d(x_n, x_{n_1}) < \frac{1}{2} \quad \text{for all } n \geq n_1.$$

Again since $\langle x_n \rangle$ is a Cauchy sequence, there exists an $n_2 \in \mathbb{Z}_+$ such that $n_2 > n_1$, and

$$d(x_n, x_{n_2}) < \frac{1}{2^2} \quad \text{for all } n \geq n_2.$$

Continuing in this way, there exists an $n_k \in \mathbb{Z}_+$ such that $n_k > n_{k-1}$, and

$$d(x_n, x_{n_k}) < \frac{1}{2^k} \quad \text{for all } n \geq n_k.$$

Let us denote by F_k the closure in X of the sphere $S\left(x_{n_k}, \frac{1}{2^{k-1}}\right)$, that is,

$$F_k = \overline{S\left(x_{n_k}, \frac{1}{2^{k-1}}\right)}.$$

Each F_k is non-empty since $x_{n_k} \in F_k$. We show that $\langle F_k \rangle$ is a nested sequence. Let y be a point in F_{k+1}. Then $d(x_{n_{k+1}}, y) \leq \frac{1}{2^k}$. Now

$$\begin{aligned} d(x_{n_k}, y) &\leq d(x_{n_k}, x_{n_{k+1}}) + d(x_{n_{k+1}}, y) \\ &< \frac{1}{2^k} + \frac{1}{2^k} = \frac{1}{2^{k-1}}. \end{aligned}$$

Therefore, y is in $S(x_{n_k}, \frac{1}{2^{k-1}})$ which is contained in F_k. Also F_k, being the closure of a set, is a closed set in X. Further, the diameter of F_k

$$\operatorname{diam} F_k \leq \frac{2}{2^{k-1}}$$

which tends to zero as k tends to infinity. Thus, $\langle F_k \rangle$ is a nested sequence of non-empty closed sets such that $\operatorname{diam} F_k$ tends to zero as k tends to infinity. Therefore, by our assumption, $\bigcap_{k=1}^{\infty} F_k$ contains exactly one point $x \in X$. Since x is in F_k for each k,

$$d(x_{n_k}, x) \leq \frac{1}{2^{k-1}}.$$

This means that the subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ converges to x in X. This, in turn, means that the sequence $\langle x_n \rangle$ converges to x, and the proof is complete.

The following examples show that no condition in the Cantor's intersection theorem can be dropped without losing the conclusion of the theorem.

Example 5.3.3 Let $\mathbb{R}$ be the set of real numbers, and d be the discrete metric on X. $(\mathbb{R}, d)$ is a complete metric space. For each positive integer n, let us define

$$F_n = \{x \in \mathbb{R} : x > n\}.$$

Then $\langle F_n \rangle$ is a nested sequence of non-empty closed sets in $(\mathbb{R}, d)$. The $\operatorname{diam} F_n = 1$ for each n. The intersection $\bigcap_{n=1}^{\infty} F_n$ is empty.

Example 5.3.4 Let $\mathbb{Q}$ be the set of rational numbers, and d be the usual metric on $\mathbb{Q}$. We have already shown that $(\mathbb{Q}, d)$ is not complete. Let

$$F_n = \{x \in \mathbb{Q}_+ : 2 < x^2 \leq 2 + \frac{1}{n}\} \quad \text{for all } n \in \mathbb{Z}_+$$

where $\mathbb{Q}_+$ is the set of positive rational numbers. Then $\langle F_n \rangle$ is a sequence of non-empty closed sets in F with diameter tending to zero. Also the intersection $\bigcap_{n=1}^{\infty} F_n$ is empty.

Example 5.3.5 Consider the two dimensional Euclidean plane $\mathbb{R}^2$. Let

$$F_n = S((0,0), \frac{1}{n}) - \{(0,0)\} \quad \text{for all } n \in \mathbb{Z}_+.$$

Then $\langle F_n \rangle$ is a nested sequence of non-empty sets with $\operatorname{diam} F_n$ tending to zero as n tends to infinity. Here F_n's are not closed sets. The point $(0,0)$ is a limit point of F_n, and it is not in F_n. It can be seen that $\bigcap_{n=1}^{\infty} F_n$ is empty.

EXERCISE

1. Prove that a metric space is complete if and only if the intersection of every nested sequence of non-empty closed spheres with radii tending to zero is non-empty.

2. Let $\mathbb{N}$ be equipped with the metric

$$d(m,n) = \begin{cases} 0 & \text{if } m = n, \\ 1 + \frac{1}{m+n} & \text{if } m \neq n. \end{cases}$$

   Show that $(\mathbb{N}, d)$ is a complete metric space. Define $S_n = \{m : m \geq n\}$. Prove that $\langle S_n \rangle$ is a nested sequence of closed spheres whose intersection is empty.

3. Show that the metric space in Example 2.1.24 is complete, and that it is a closed subspace of the metric space in Example 2.1.23.

5.4 Completion of a Metric Space

A metric space (X, d) is said to be embedded in a metric space (X^*, d^*) if there is an isometry from X to a subset Y of X^*. If Y is dense in X^*, and X^* is complete then (X^*, d^*) is called completion of (X, d). The subspace Y of (X^*, d^*) above, for all analytical purposes, has the same properties as the given metric space (X, d), and thus we identify the subspace Y with the metric space (X, d).

Theorem 5.4.1 Every metric space is a dense subspace of a complete metric space.

Proof. Let (X, d) be a metric space, and X_1 be the set of all Cauchy sequences in X. We define a relation on X_1 as follows:

$$\langle x_n \rangle \sim \langle y_n \rangle \text{ if } \lim_{n\to\infty} d(x_n, y_n) = 0 \quad \text{for all } \langle x_n \rangle \text{ and } \langle y_n \rangle \text{ in } X_1.$$

Here the limit $\lim_{n\to\infty} d(x_n, y_n)$ is taken in the real line $\mathbb{R}$. The relation $\sim$ is an equivalence relation on X_1. The reflexivity and symmetry of $\sim$ are obvious. For the transitivity of the relation $\sim$, assume that $\langle x_n \rangle \sim \langle y_n \rangle$ and $\langle y_n \rangle \sim \langle z_n \rangle$. Then, taking limits as n tends to infinity in the following inequality

$$d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n)$$

it follows that $\langle x_n \rangle \sim \langle z_n \rangle$. Consequently, the relation $\sim$ partitions X_1 into equivalence classes. Let $[\langle x_n \rangle]$ denote the equivalence class of $\langle x_n \rangle$ in X_1, and define

$$Y = \{[\langle x_n \rangle] : \langle x_n \rangle \in X_1\}.$$

Next we convert Y into a metric space. Let $d' : Y \times Y \to \mathbb{R}$ be a mapping defined by

$$d'(x, y) = \lim_{n\to\infty} d(x_n, y_n) \quad \text{for all } x = [\langle x_n \rangle] \text{ and } y = [\langle y_n \rangle] \text{ in } Y.$$

First we show that that d' is well-defined. For this we need to ensure that the limit on the right in the definition of d', $\lim_{n\to\infty} d(x_n, y_n)$, actually exists, and that $\lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} d(x'_n, y'_n)$ for all sequences $\langle x'_n\rangle$ in x and $\langle y'_n\rangle$ in y. Let $\epsilon > 0$ be a real number. Since $\langle x_n\rangle$ and $\langle y_n\rangle$ are Cauchy sequences, there exists positive integers p_1 and p_2 such that

$$d(x_n, x_m) < \frac{\epsilon}{2} \quad \text{for all } m, n \geq p_1$$

and

$$d(y_n, y_m) < \frac{\epsilon}{2} \quad \text{for all } m, n \geq p_2.$$

Also by Lemma 3.1.6

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m).$$

Let $p_0 = \max\{p_1, p_2\}$. Then

$$|d(x_n, y_n) - d(x_m, y_m)| < \epsilon \quad \text{for all } n, m \geq p_0.$$

This proves that $\langle d(x_n, y_n)\rangle$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $\lim_{n\to\infty} d(x_n, y_n)$ exists.

Now let $\langle x'_n\rangle$ be a sequence in x and $\langle y'_n\rangle$ be a sequence in y. The limits $\lim_{n\to\infty} d(x_n, y_n)$, $\lim_{n\to\infty} d(x_n, y'_n)$, $\lim_{n\to\infty} d(x'_n, y_n)$ and $\lim_{n\to\infty} d(x'_n, y'_n)$ exist. Now taking the limits as n tends to infinity in the inequality

$$d(x_n, y_n) \leq d(x_n, x'_n) + d(x'_n, y'_n) + d(y'_n, y_n)$$

it follows that

$$\lim_{n\to\infty} d(x_n, y_n) \leq \lim_{n\to\infty} d(x'_n, y'_n).$$

Similarly, $\lim_{n\to\infty} d(x'_n, y'_n) \leq \lim_{n\to\infty} d(x_n, y_n)$. Thus, $\lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} d(x'_n, y'_n)$, and d' is well-defined.

Next we show that (Y, d') is a metric space. Since $\lim_{n\to\infty} d(x_n, y_n) \geq 0$ for all $\langle x_n\rangle$ and $\langle y_n\rangle$ in x and y, respectively, $d'(x, y) \geq 0$. Now $d'(x, y) = 0$ if and only if $\lim_{n\to\infty} d(x_n, y_n) = 0$, for all $\langle x_n\rangle$ in x and $\langle y_n\rangle$ in y if and only if $\langle x_n\rangle \sim \langle y_n\rangle$ if and only if $x = [\langle x_n\rangle] = [\langle y_n\rangle] = y$. The symmetry for d' is obvious. Now let $x = [\langle x_n\rangle]$, $y = [\langle y_n\rangle]$ and $z = [\langle z_n\rangle]$ be any three elements in Y. Now taking limits as n tends to infinity in the triangle inequality

$$d(x_n, y_n) \leq d(x_n, z_n) + d(z_n, y_n)$$

we get $d'(x, y) \leq d'(x, z) + d'(z, y)$. Thus, (Y, d') is a metric space.

Next we show that (X, d) is isometric to a subspace of (Y, d'). A sequence x_n in X is called stationary if it is a constant sequence, that is, $x_n = x_m$ for all m and n. We define the set

$$Y_0 = \{x \in Y : \text{ there exists a stationary sequence } \langle x_n\rangle \text{ in } x\}.$$

Since a stationary sequence is a Cauchy sequence, each stationary sequence is in some equivalence class in Y. Now, let $\langle x_n\rangle$ and $\langle y_n\rangle$ be any two distinct stationary sequences in X. Then $\lim_{n\to\infty} d(x_n, y_n) = d(x_1, y_1) \neq 0$. This means that $\langle x_n\rangle$ is not related to $\langle y_n\rangle$. Thus, an equivalence class in Y contains at the most one stationary sequence. We define a mapping $f : X \to Y_o$ by setting $f(x) = y$ if the stationary sequence $\langle x_n\rangle$, where $x_n = x$ for all n, is in y. The mapping f is surjective since for each $y \in Y_0$, there is a point x in X such that the stationary sequence $\langle x_n\rangle$, where $x_n = x$ for all n, is in y, and, by the definition,

$f(x) = y$. To show that f is an isometry, let x and x' be any two elements in X, and $\langle x_n \rangle$ and $\langle x'_n \rangle$ be the stationary sequences such that $x_n = x$ and $x'_n = x'$ for all n. Then $\langle x_n \rangle$ is in $f(x)$, $\langle x'_n \rangle$ is in $f(x')$, and

$$d'(f(x), f(x')) = \lim_{n\to\infty} d(x_n, x'_n) = d(x, x').$$

Next we show that Y_0 is dense subset of Y. To prove this, let $\epsilon > 0$ be a real number. Let y be any element in $Y - Y_0$. Then there exists a sequence $\langle x_n \rangle$ in X_1 such that $[\langle x_n \rangle] = y$. Since $\langle x_n \rangle$ is a Cauchy sequence, there exists an $n_0 \in \mathbb{Z}_+$ such that

$$d(x_n, x_{n_0}) < \frac{\epsilon}{2} \quad \text{for all} \quad n \geq n_0.$$

Let $f(x_{n_0}) = y_0$. Then the stationary sequence $\langle z_n \rangle$, where $z_n = x_{n_0}$ for all n, is in y_0. Therefore,

$$d'(y, y_0) = \lim_{n\to\infty} d(x_n, x_{n_0}) \leq \frac{\epsilon}{2} < \epsilon.$$

Hence, y is a limit point of Y_0, and so $\bar{Y_0} = Y$.

It remains to show that (Y, d') is complete. Let $\langle y_n \rangle$ be a Cauchy sequence in Y. Since Y_0 is dense in Y, for each positive integer n, there exists a point z_n in Y_0 such that

$$d'(y_n, z_n) < \frac{1}{n}.$$

Since $f : X \to Y_0$ is an isometry, there is a point x_n in X such that $f(x_n) = z_n$, and that the stationary sequence $\langle x_k^n \rangle$ such that $x_k^n = x_n$ for all $k \in \mathbb{Z}_+$ is in z_n.
Now

$$\begin{aligned} d(x_n, x_m) &= d'(f(x_n), f(x_m)) = d'(z_n, z_m) \\ &\leq d'(z_n, y_n) + d'(y_n, y_m) + d'(y_m, z_m) \\ &< \frac{1}{n} + d'(y_n, y_m) + \frac{1}{m}. \end{aligned}$$

Since $\langle y_n \rangle$ is a Cauchy sequence, there exists a $p \in \mathbb{Z}_+$ such that $d'(y_n, y_m) < \frac{\epsilon}{2}$ for all $n, m \geq p$. Also, there is a $p' \in \mathbb{Z}_+$ such that $\frac{1}{n} + \frac{1}{m} < \frac{\epsilon}{2}$ for all $n, m \geq p'$. Let $p_0 = \max\{p, p'\}$. Then

$$d(x_n, x_m) < \epsilon \quad \text{for all} \quad n, m \geq p_0.$$

This shows that $\langle x_n \rangle$ is a Cauchy sequence in X. Therefore, there exists a point y in Y such that $\langle x_n \rangle$ is in y. Now

$$d'(y_n, y) \leq d'(y_n, z_n) + d'(z_n, y) < \frac{1}{n} + d'(z_n, y).$$

Also,

$$d'(z_n, y) = \lim_{k\to\infty} d(x_k^n, x_k) = \lim_{k\to\infty} d(x_n, x_k).$$

Since $\langle x_n \rangle$ is a Cauchy sequence, there exists a $q \in \mathbb{Z}_+$ such that

$$d(x_n, x_k) < \frac{\epsilon}{2} \quad \text{for all} \quad n, k \geq q.$$

This shows that $\lim_{k\to\infty} d(x_n, x_k) \leq \frac{\epsilon}{2}$ for all $n \geq q$. Also there exists a $q' \in \mathbb{Z}_+$ such that $\frac{1}{n} < \epsilon$ for all $n \geq q'$. Let $q_0 = \max\{q, q'\}$. Then

$$d'(y_n, y) < \epsilon \quad \text{for all} \quad n \geq q_0.$$

So the sequence $\langle y_n \rangle$ converges to y, and (Y, d') is complete. This completes the proof.

As the proof of the above theorem is a bit lengthy, the following outline of the proof is in order.

1. Introduce an equivalence relation $\sim$ on the set X_1 of all Cauchy sequences in X by putting
$$\langle x_n \rangle \sim \langle y_n \rangle \text{ if } \lim_{n\to\infty} d(x_n, y_n) = 0.$$

2. Denote by Y the set of all equivalence classes of the relation $\sim$, and show that (Y, d') is a metric space with the metric
$$d'([\langle x_n \rangle], [\langle x'_n \rangle]) = \lim_{n\to\infty} d(x_n, x'_n),$$
where $[\langle x_n \rangle]$ is the equivalence class of $\langle x_n \rangle$ in X_1.

3. Denote by Y_0 the set of points in Y containing stationary sequences. Show that $f : X \to Y_0$ defined by $f(x) = y$, where y contains the stationary sequence $\langle x_n \rangle$ such that $x_n = x$ for all n, is an isometry.

4. Show that Y_0 is dense in Y.

5. Show that (Y, d') is complete.

The reader shall notice that the above proof runs parallel to the proof of completion of rationals (see Chapter 10). Infact the motivation for the above proof of completion of metric spaces originally came from the proof of completion of rational numbers.

The next theorem shows that completion of a metric space is unique upto isometries. That is, if (Y_1, d_1) and (Y_2, d_2) are two completions of a metric space (X, d), then (Y_1, d_1) and (Y_2, d_2) are isometric to each other.

Theorem 5.4.2 Completion of a metric space is unique upto isometries.

Proof. Let (X, d) be a metric space, and (Y, d_1) and (Z, d_2) be two completions of (X, d). Also, assume that X is dense subspace of both Y and Z.

We define an isometry from Y to Z. Let y be a point in Y. Since X is dense in Y, there exists a sequence $\langle x_n \rangle$ in x which, as a sequence in Y converges to y. Then $\langle x_n \rangle$ is a Cauchy sequence in X. Since $X \subseteq Z$, $\langle x_n \rangle$ is a Cauchy sequence in Z. Since Z is complete, $\langle x_n \rangle$ converges to a point z in Z . We define a mapping $f : Y \to Z$ by setting $f(y) = z$.

First we show that f is well-defined. For this we need to show that if $\langle x'_n \rangle$ is a sequence in X which, as a sequence in Y converges to y, then $\langle x'_n \rangle$ as a sequence in Z converges to z. Suppose that $\langle x'_n \rangle$ as a sequence in Z converges to a point z' in Z. Now

$$d_2(z, z') \leq d_2(z', x'_n) + d_2(x'_n, x_n) + d_2(x_n, z). \tag{A}$$

Let $\epsilon > 0$ be a real number. Since $\langle x'_n \rangle$ converges to z' in Z, there exists an $m_1 \in \mathbb{Z}_+$ such that

$$d_2(z', x'_n) < \frac{\epsilon}{3} \quad \text{for all} \quad n \geq m_1.$$

Since both $\langle x_n \rangle$ and $\langle x'_n \rangle$ converge to y in Y, by the inequality $d_1(x_n, x'_n) \leq d_1(x_n, y) + d_1(y, x'_n)$, it follows that there is an $m_2 \in \mathbb{Z}_+$ such that

$$d_1(x_n, x'_n) < \frac{\epsilon}{3} \quad \text{for all} \quad n \geq m_2.$$

Therefore, $d_2(x_n, x'_n) = d(x_n, x'_n) = d_1(x_n, x'_n) < \frac{\epsilon}{3}$ for all $n \geq m_2$. Again since $\langle x_n \rangle$ converges to z in Z, there exists an $m_3 \in \mathbb{Z}_+$ such that

$$d_2(x_n, z) < \frac{\epsilon}{3} \quad \text{for all} \quad n \geq m_3.$$

Let $m_0 = \max\{m_1, m_2, m_3\}$. Then for $n \geq m_0$, the statement,(A) proves that

$$d_2(z, z') < \epsilon.$$

Since ϵ is arbitrary, this shows that $z = z'$, and f is well-defined.

Now we show that f is surjective. Let z be any element in Z. Then there exists a sequence $\langle x_n \rangle$ in X such that $\langle x_n \rangle$ as a sequence in Z converges to z. Then $\langle x_n \rangle$ is a Cauchy sequence in X, and so it is a Cauchy sequence in Y. By the completeness of Y, the sequence $\langle x_n \rangle$ will converge to a point y in Y. Then, by our definition of f, $f(y) = z$.

Now it remains to show that f is distance preserving. Let y and y' be two elements in Y, and let z and z' be two elements in Z such that $f(y) = z$, and $f(y') = z'$. Then there exist sequences $\langle x_n \rangle$ and $\langle x'_n \rangle$ in X such that $\langle x_n \rangle$ converges to y in Y, $\langle x_n \rangle$ converges to z in Z, $\langle x'_n \rangle$ converges to y' in Y, and x'_n converges to z' in Z. Now, by the triangle inequality

$$d_1(y, y') \leq d_1(y, x_n) + d_1(x_n, x'_n) + d_1(x'_n, y'). \tag{B}$$

Let $\epsilon > 0$ be a real number. Since $\langle x_n \rangle$ converges to y in Y, and $\langle x'_n \rangle$ converges to y' in Y, there exists a $p \in \mathbb{Z}_+$ such that

$$d_1(y, x_n) < \frac{\epsilon}{2} \text{for all } n \geq p,$$

$$d_1(x'_n, y') < \frac{\epsilon}{2} \text{for all } n \geq p.$$

Therefore, $\lim_{n\to\infty} d_1(y, x_n) \leq \frac{\epsilon}{2}$, and $\lim_{n\to\infty} d_1(x'_n, y') \leq \frac{\epsilon}{2}$. Since ϵ is arbitrary, $\lim_{n\to\infty} d_1(y, x_n) = 0$, and $\lim_{n\to\infty} d_1(x'_n, y') = 0$. Now taking the limit in statement (B) as n tends to infinity, we have

$$d_1(y, y') \leq \lim_{n\to\infty} d_1(x_n, x'_n).$$

Interchanging x_n with y and x'_n with y' in (B), and taking limits as n tends to infinity, we have

$$\lim_{n\to\infty} d_1(x_n, x'_n) \leq d_1(y, y').$$

Thus, $d_1(y, y) = \lim_{n\to\infty} d_1(x_n, x'_n)$. Similarly,

$$d_2(z, z') = \lim_{n\to\infty} d_2(x_n, x'_n) = \lim_{n\to\infty} d_1(x_n, x'_n).$$

Therefore, $d_1(y, y') = d_2(z, z') = d_2(f(y), f(y'))$, and the proof is complete.

Since completion of a metric space is unique upto isometry, we use the word "the" completion of a metric space in this sense.

We have seen in Theorem 5.4.1 that there is an isometry f between (X, d) and a dense subsapce Y_0 of the complete metric space (Y, d'). We identify an element x in X with its image $f(x)$ in Y_0. Once this is done the metric space X can be considered literally as a dense subspace of its completion, and the isometry f is replaced by the inclusion mapping of X into Y, and the following definition thus makes sense which is helpful in obtaining completions of several examples of metric spaces.

Definition 5.4.3 A metric space (X^*, d^*) is called a completion of a metric space (X, d) if

(i) (X^*, d^*) is complete, and

(ii) (X, d) is a dense subspace of (X^*, d^*).

Examples 5.4.4

1. The real line $\mathbb{R}$ is completion of itself. Note that every metric space is a dense subspace of itself.

2. A completion of the space of rationals $\mathbb{Q}$ with usual metric is the real $\mathbb{R}$ since $\mathbb{Q}$ is dense in $\mathbb{R}$, and $\mathbb{R}$ is complete.

3. The completion of $]a, \infty[$, $]-\infty, b[$, $a, b \in \mathbb{R}$ are $[a, \infty[$ and $]-\infty, b]$ respectively. The completion of $]a, b[$ is $[a, b]$.

4. Let (X, d) be discrete metric space, and Y be a subspace of X. The completion of Y is Y itself.

5. In the two dimensional Euclidean plane $\mathbb{R}^2$, the completion of the subspace

$$Y = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$$

is the subspace $Y^* = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$ of $\mathbb{R}^2$.

6. The completion of the subspace $\mathbb{Q} \times \mathbb{Q}$ of two dimensioanl Ecuclidean plane $\mathbb{R}^2$ is $\mathbb{R}^2$.

7. The completion of the subspace $S = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}$ in n-dimensional Euclidean plane $\mathbb{R}^n$ is S itself.

8. Let S be an open sphere in l_2-space. Then the completion of S is $\bar{S}$, the closure S in l_2-space.

EXERCISE

1. Show that the completion of a totally bounded metric space is totally bounded.
2. Let Y be an everywhere dense subset of a metric space X. If every Cauchy sequence in Y converges in X, then X is complete.
3. A metric space is separable if and only if its completion is separable.
4. Let (X, d) be a complete metric space. Show that there exists a uniformly continuous mapping $f : X \to]0, \infty[$ such that the infimum $\inf_{x \in X} f(x) = 0$. .

5.5 Baire's Category Theorem

Intutively, the notions of dense set and nowhere dense set classify subsets of a metric space in terms of the scatteredness of their elements in the entire metric space. A set A is dense in a metric space X if the elements of A are arbitrarily close to each element of X. Similarly, a set A is nowhere dense in a metric space X if the elements of $X - \bar{A}$ are arbitrarily close to each element of A, that is, for each element x of $\bar{A}$, there is no neighbourhood of x consisting entirely of the points of $\bar{A}$. For example, the set $\mathbb{Q}$ is dense in the real line $\mathbb{R}$, and the set $\mathbb{Z}$ of integers is nowhere dense in $\mathbb{R}$; $\mathbb{Z}$ does not cover much of $\mathbb{R}$. We have already shown that countably infinite union of nowhere dense sets need not be a nowhere dense set. It is the genius of French mathematician René Baire that recognised the fundamental role played by the sets that can be written as the union of countable number of nowhere dense sets in proving several existence theorems in pure mathematics.

Definition 5.5.1 A subset of a metric space X is said to be of first category in X if it is the union of a countable collection of nowhere dense sets in X. A set is said to be of second category in X if it is not of first category in X.

Examples 5.5.2

1. The subset $\mathbb{Q} \times \mathbb{Q}$ of $\mathbb{R}^2$ is of first category.
2. Every non-empty subset in a discrete metric space is of second category.
3. The subset of rationals $\mathbb{Q}$ in $\mathbb{R}$ is of first category in $\mathbb{R}$.
4. Every nowhere dense set in a metric space X is of first category in X.

Theorem 5.5.3 Let (X, d) be a metric space, and (Y, d) be a subspace of X. If a subset A of Y is of first category in Y, then A is of first category in X. Dually, if a subset A of Y is of second category in X, then A is of second category in Y.

Proof. Assume that $A \subseteq Y$, and A is not nowhere dense set in X. Let $B = (\bar{A})^0$. Then B is non-empty, and $B \subseteq \bar{A} \subseteq \bar{Y}$. We show that $B \cap Y \neq \emptyset$. Suppose that $B \cap Y = \emptyset$. Let x be a point in B. Since B is open, there exists a real number $\epsilon > 0$ such that $S(x, \epsilon) \subseteq B$. This means that $S(x, \epsilon) \cap Y = \emptyset$. Again since $B \subseteq \bar{Y}$, $x \in \bar{Y} - Y$ so that x is a limit point of Y.

But this is a contradiction. Therefore, $B \cap Y$ is non-empty. Since B is open in X, $B \cap Y$ is open in Y. Also $B \cap Y \subseteq \bar{A} \cap Y$. This proves that the closure of A in Y has non-empty inetrior. Therefore, A is not nowhere dense in Y. The proof now follows.

Corollary 5.5.4 If a metric space X is of first category in itself, then X is of first category in any metric space containing X as a subspace. If a subset Y of X is of second category in X, then Y is of second category in Y.

Theorem 5.5.5 Let A be a subset of second category in a metric space X, and $\{M_n\}_{n=1}^{\infty}$ be a countable covering of A consisting of subsets of X. Then $A \cap (\bar{M}_n)^0 \neq \emptyset$ for at least one $n \in \mathbb{Z}_+$.

Proof. Suppose on the contrary that $A \cap (\bar{M}_n)^0 = \emptyset$ for all $n \in \mathbb{Z}_+$. Then, $A \subseteq \bigcup_{n=1}^{\infty}((\bar{M})_n - (\bar{M}_n)^0)$, where for each n, $\bar{M}_n - (\bar{M}_n)^0$ is nowhere dense set in X.

Theorem 5.5.6 (Baire's category). Every non-empty open set in a complete metric space is of second category.

Proof. Let X be a complete metric space, and G be a non-empty open set in X. Let $\langle O_n\rangle_{n=1}^{\infty}$ be a sequence of nowhere dense sets in X. We show that $G \neq A = \bigcup_{n=1}^{\infty} O_n$. By Theorem 2.11.8, there exists an open sphere, $S_1 = S(x_1, \epsilon_1)$, where the radius ϵ_1 is a real number such that $0 < \epsilon_1 < 1$, such that $\bar{S}_1 \subseteq G$, and $\bar{S}_1 \cap O_1 = \emptyset$. By induction, we can construct a sequence $\langle S_n\rangle$ for which $S_{n+1} \subseteq S_n$, $\epsilon_n < \frac{1}{n}$, $\bar{S}_n \subseteq S_{n-1}$, and $\bar{S}_n \cap O_n = \emptyset$ for all $n \geq 2$. The sequence $\langle x_n\rangle$ of centers x_n in S_n is then a Cauchy sequence in X. Since X is complete, the sequence $\langle x_n\rangle$ converges to a point x in X. Also note that the subsequence $\langle x_{n+1}, x_{n+2}, \dots\rangle$ of $\langle x_n\rangle$ is in S_{n+1}, and converges to x. Therefore, $x \in \bar{S}_{n+1} \subseteq S_n$ for each $n \in \mathbb{Z}_+$. Thus, x is not in A. But $x \in \bar{S}_1 \subseteq G$. This completes the proof.

Corollary 5.5.7 A complete metric space is of second category.

Theorem 5.5.8 In a complete metric space, the intersection of a countable family of dense open sets is dense.

Proof. Let $\{O_i\}_{i\in\mathbb{N}}$ be a countable family of dense open sets in a complete metric space (X, d). We need to show that $\bigcap_{i\in\mathbb{N}} O_i$, $i \in \mathbb{N}$ is dense in X. Let x_1 be any point in X, and $r_1 > 0$ be a real number. We show that $\bigcap_{i\in\mathbb{N}} O_i$ intersects $S_1 = S(x_1, r_1)$. Since O_1 is open and dense in X, $O_1 \cap S_1$ is open and non-empty. Then there exists a point x_2 in $O_1 \cap S_1$ and a real number $r_2 > 0$ such that the closed sphere $S_2' = S[x_2, r_2]$ is contained in $O_1 \cap S_1$. Now consider the open sphere $S_2 = S(x_2, r_2)$. Since O_2 is open and dense in X, $O_2 \cap S_2$ is non-empty open set. So there exists a point x_3 in $O_2 \cap S_2$ and a real $r_3 > 0$ such that the closed sphere $S_3' = S[x_3, r_3]$ is contained in $O_2 \cap S_2$. Continuing in this way, by induction, we get

$$S'_{n+1} \subseteq O_n \cap S_n,$$

moreover, we can insist that r_n tends to zero as n tends to infinity. We get a sequence $\langle x_n\rangle$ such that $x_n \in S_n'$. If i, $j > n$, then

$$d(x_i, x_j) \leq d(x_i, x_n) + d(x_n, x_j) < 2r_n.$$

This means that $\langle x_n\rangle$ is a Cauchy sequence. Since X is complete, there exists a point $x_0 \in X$ such that $\langle x_n\rangle$ converges to x_0. Since for $i > n$,

$$x_i \in S'_{n+1} \subseteq S_1 \cap O_n.$$

We can let i tend towards infinity to conclude that $x_0 \in S'_{n+1} \subseteq S_1 \cap O_n$. Since this is true for all n, the set $\bigcap_{n=1}^{\infty} O_n$ does indeed intersect S_1.

Corollary 5.5.9 If a complete metric space is expressed as a countable union of closed sets, then one of the closed sets must have a non-empty interior.

Proof. Let (X, d) be a complete metric space, and let $X = \bigcup_{n=1}^{\infty} F_n$, where each F_n is a closed set in X having empty interior. The sets $O_n = X - F_n$ are non-empty open and dense in X. Hence, by Baire's theorem, $\bigcap_{n=1}^{\infty} O_n$ is dense in X. In particular, it is non-empty. If $x \in \bigcap_{n=1}^{\infty} O_n$, then $x \in X - \bigcup_{n=1}^{\infty} F_n$, a contradiction.

The following corollary is an immediate consequence of the above corollary.

Corollary 5.5.10 If a complete metric space X is expressed as a countable union of subsets of X, then the closure of one of the sets must have non-empty interior.

Example 5.5.11 The set of irrational numbers $\mathbb{R} - \mathbb{Q}$ in the real line is of second category. Suppose on the contrary that this set is of first category. Since the set of rational numbers is of first category, the union $(\mathbb{R} - \mathbb{Q}) \cup \mathbb{Q} = \mathbb{R}$ is of first category. This contradicts the fact that the real line $\mathbb{R}$, being a complete metric space, is of second category by Baire's Category theorem.

There are many applications of Baire's theorem in analysis and most of them are outside the scope of this book. Here we give two applications of it to the theory of real functions of real variables. Recall the function given by

$$f(x) = \begin{cases} \dfrac{1}{q} & \text{if } x = \dfrac{p}{q} \text{ in reduced form,} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

which is a standard example of a function that is continuous at each irrational point of $]0, 1[$ and discontinuous at each rational point of $]0, 1[$. The graph of this function is shown in Fig. 5.2.

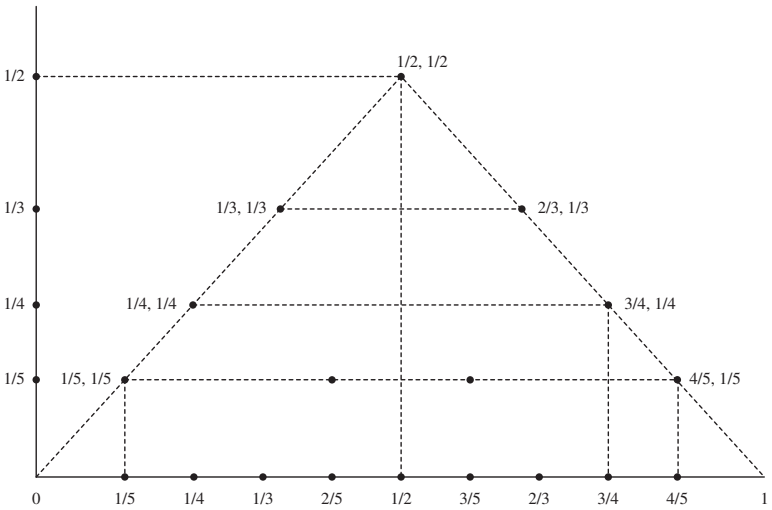


Fig. 5.2

It is now natural to ask whether there is a function defined on $]0,1[$ continuous at each rational point of $]0,1[$ and discontinuous at each irrational point of $]0,1[$. We next prove that there is no such function, and this makes use of Baire's theorem.

Recall from elementary calculus that for any bounded real valued function f defined on an open interval $I =]a,b[$, the oscillation of f on I is given by

$$\omega_f(I) = \sup\{f(x) : x \in I\} - \inf\{f(x) : x \in I\},$$

and the oscillation of f at a point a in I is given by

$$\omega_f(a) = \inf\{\omega_f(J) : J \text{ is an open interval such that } a \in J \subseteq I\}.$$

The following lemma relates continuity and oscillation of a function.

Lemma 5.5.12 A bounded real valued function defined on an open interval I is continuous at a point $a \in I$ if and only $\omega_f(a) = 0$, and the set $\{x \in I : \omega_f(x) < \epsilon\}$is open for each real number $\epsilon > 0$.

Proof. Let f be continuous at a. Then for each real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that

$$|x - a| < \delta \text{ implies that } |f(x) - f(a)| < \epsilon.$$

Therefore,

$$\omega_f(a) \leq \omega_f(]a - \delta, a + \delta[) < \epsilon.$$

Since ϵ is arbitrary, $\omega_f(a) = 0$. Conversely, let $\omega_f(a) = 0$. Then for any real number $\epsilon > 0$, there exists an open interval $J \subseteq I$ such that $a \in J$ and $\omega_f(J) < \epsilon$. Since J is open, there exists a real number $\delta > 0$ such that $]a - \delta, a + \delta[\subseteq J$. Therefore, $\omega_f(]a - \delta, a + \delta[) < \epsilon$. Thus,

$$|x - a| < \delta \quad \text{implies that} \quad |f(x) - f(a)| \leq \omega_f(]a - \delta, a + \delta[) \leq \omega_f(J) < \epsilon.$$

To prove the second assertion of the lemma, let $\epsilon > 0$ be a real number, and x_0 be a point in the set $\{x \in I : \omega_f(x) < \epsilon\}$. Therefore, there exists an open interval J containing x_0 and contained in I such that $\omega_f(J) < \epsilon$. Now for any $y \in J$,

$$\omega_f(y) \leq \omega_f(J) < \epsilon.$$

Therefore, $J \subseteq \{x \in I : \omega_f(x) < \epsilon\}$, and the proof is complete.

Theorem 5.5.13 There is no function $f :]0,1[\to \mathbb{R}$ that is continuous at each rational point of $]0,1[$ and discontinuous at each irrational point in $]0,1[$.

Proof. Suppose on the contrary that there is a function $f :]0,1[\to \mathbb{R}$ which is continuous at each rational point of $]0,1[$ and discontinuous at each irrational point of $]0,1[$. By the above lemma, for each $n \in \mathbb{Z}_+$, the set

$$U_n = \{x \in]0,1[: \omega_f(x) < \frac{1}{n}\}$$

is open. Again using the first part of the above lemma, the set $\bigcap_{n=1}^{\infty} U_n$ is equal to $\mathbb{Q} \cap]0,1[$. Since the rational numbers in $]0,1[$ are dense in $]0,1[$, $\bigcap_{n=1}^{\infty} U_n$, is dense in $]0,1[$. Therefore,

each set U_n is dense in $]0,1[$. Let $V_n =]0,1[- U_n$ for each $n \in \mathbb{Z}_+$. Then $\bigcup_{n=1}^{\infty} V_n =]0,1[- \bigcap_{n=1}^{\infty} U_n$, and each V_n is nowhere dense in $]0,1[$. If $r_1, r_2, \ldots$ is an enumeration of rationals in $]0,1[$, then

$$]0,1[= V_1 \cup \{r_1\} \cup V_2 \cup \{r_2\} \cup \ldots$$

and so

$$[0,1] = \{0\} \cup \{1\} \cup V_1 \cup \{r_1\} \cup V_2 \cup \{r_2\} \cup \ldots \ .$$

Each set in this union is nowhere dense, and thus $[0,1]$ is of first category. This contradicts Baire's category theorem, since $[0,1]$ is a complete metric space, and the proof is complete.

The second application of Baire's theorem will be discussed in Section 8.5.

EXERCISE

1. Is every set of the second category the complement of a set of first category?
2. Prove that in a complete metric space, the complement of a set of the first category is dense, and is of second category.
3. Give an example of a metric space Z, a subspace X of Z and a subset E of X such that E is of second category in X, and is of first category in Z. This shows that the concepts of first category and second category depend on the metric spaces.
4. Every point of a dense-in-itself complete metric space is a condensation point of the space.
5. Every denumerable complete metric space has infinite number of isolated points.
6. Determine the category of these sets:
   (a) the union of all vertical lines in $\mathbb{R}^2$ that pass through a rational point on the horizontal axis,
   (b) the set of all polynomials in $\mathcal{C}[a,b]$.
7. Does a homeomorphism preserve the category of sets?
8. Is the closure of a nowhere dense set also nowhere dense? Is the closure of a set of first category also of first category?
9. Prove that the set of real numbers whose decimal expansions do not contain the digit 7 is of first category.
10. Select a function $f_0 \in \mathcal{C}[0,1]$ and a sequence of real numbers $\langle \alpha_n \rangle$. Define recursively

$$f_{n+1}(t) = \alpha_n + \int_0^t f_n(x)dx, n = 0, 1, 2, \ldots .$$

    Assume that for each $x \in [0,1]$, there is an n for which $f_n(x) = 0$. Prove that $f_0 = 0$.
11. Prove that in a complete metric space, a set of first category has empty interior.

12. If a complete metric space is the union of a countable number of closed sets, then at least one of them must contain a non-empty open set.

13. Give an example of a metric space having countably many points that contains no subset of second category.

14. Prove the following theorem:

    Let X be a complete metric space. Let $\{f_\alpha\}$ be a family of continuous real-valued mappings defined on X. Assume that for each x, $\sup_\alpha |f_\alpha(x)| < \infty$. Then for some non-empty open set G, $\sup_{x\in G} |f_\alpha(x)| < \infty$.

15. Prove that every subset of first category set is a first category set. Give an example of an uncountable first category set. Also prove that a countable union of first category sets is also a set of first category.

16. Is a countable set in a metric space necessarily a set of first category?

17. In a complete metric space, is every non-empty open set of second category?

18. Give an example of sets $S \subseteq F \subseteq X$, where X is a complete metric space, F is a closed set in X, and S is of second category in F but of first category in X.

19. (i) The intersection of a countable family of open sets is called a G_δ-set. Prove that the set of rationals is not a G_δ-set in the real line.

    (ii) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that each set $f^{-1}(r)$ is a G_δ-set.

    (iii) Let $f : \mathbb{R} \to \mathbb{R}$ be a mapping. Define

    $$\omega(x) = \inf_{\epsilon>0} \sup_{\substack{|x-u|<\epsilon \\ |x-v|<\epsilon}} |f(u) - f(v)|.$$

    Prove that $\omega(x) = 0$ for each x at which f is continuous. Prove that for each $\epsilon > 0$, the set $\{x : \omega(x) < \epsilon\}$ is open.

    (iv) Prove that there is no function $f : \mathbb{R} \to \mathbb{R}$ that is continuous at each rational point and discontinuous at each irrational point.

20. Prove that the complement of a nowhere dense set is dense. What about the converse?

21. Prove that a complete metric space in which there is no isolated point, is uncountable.

22. If $f : \mathbb{R} \to \mathbb{R}$, then there is an interval $]a, b[$ and a number M such that each point of $]a, b[$ is the limit of a sequence $\langle x_n \rangle$ in $]a, b[$ with $|f(x_n)| \leq M$.

23. Let X be a complete metric space, and for each n let F_n be a closed set having empty interior. Prove that $\bigcup_{n=1}^{\infty} F_n$ has empty interior.

24. Prove that a closed proper subspace in a normed linear space is nowhere dense (and hence is of first category).

25. Prove that in a complete normed linear space, subspace of second category must be dense.

26. Prove that the Cantor's set C is nowhere dense.

5.6 Completeness and Finite Intersection Property

Definition 5.6.1 Let (X, d) be a metric space, and $\mathcal{A}$ be a collection of subsets of X. Then $\mathcal{A}$ is said to have finite intersection property if every finite subcollection of $\mathcal{A}$ has non-empty intersection.

For example, the family $\{]\frac{-1}{n}, \frac{1}{n}[: n \in \mathbb{Z}_+\}$ of intervals in $\mathbb{R}$ has finite intersection property. The family of open spheres $\{S(x, \epsilon) : \epsilon$ is a positive real number$\}$ in any metric space has the finite intersection property. However, for any set X having more than one point, the family $\{\{x\} : x \in X\}$ does not have the finite intersection property.

The following theorem gives an equivalent condition for completeness.

Theorem 5.6.2 A metric space (X, d) is complete if and only if the intersection of every collection $\mathcal{A}$ of closed subsets of X which has finite intersection property, and for every real number $\epsilon > 0$, there exists a set F in $\mathcal{A}$ such that diam $(F) < \epsilon$, is non-empty.

Proof. First we assume that (X, d) is a complete metric space. Let $\mathcal{A} = \{F_i\}_{i\in\Lambda}$ be a collection of closed subsets of X with finite intersection property, and such that for each real number $\epsilon > 0$, there exists an F in $\mathcal{A}$ such that diam $(F) < \epsilon$. In particular, for each $n \in \mathbb{Z}_+$, there exists an F_{i_n} in $\mathcal{A}$ such that diam $(F_{i_n}) < \frac{1}{n}$. Then the sequence $\langle E_n\rangle$, where $E_n = \bigcap_{j=1}^{n} F_{i_j}$ satisfies all the conditions of Cantor's intersection theorem. So there is point x in X such that $\bigcap_{n=1}^{\infty} E_n = \{x\}$.
Now take arbitrary index $i_0 \in \Lambda$, and define

$$A_1 = F_{i_0},$$
$$A_{j+1} = F_{i_0} \bigcap E_j, \qquad j = 1, 2 \ldots.$$

Then the sequence $\langle A_i\rangle$ satisfies the conditions of Cantor's intersection theorem. Thus,

$$\phi \neq \bigcap_{i=1}^{\infty} A_i = F_{i_0} \bigcap \left(\bigcap_{i=1}^{\infty} E_i \right) = F_{i_0} \bigcap \{x\}.$$

Thus, x is in F_{i0} for all $i_0 \in \Lambda$, Conversely, we show that $\langle X, d\rangle$ is a complete metric space. Let $\langle F_n\rangle$ be a nested sequence of non-empty closed sets in X with diam F_n tending to zero. Then the collection of sets in this sequence $\langle F_n\rangle$ has the finite intersection property, and of course for each real number $\epsilon > 0$, there exists a set F_n in the collection such that diam $F_n < \epsilon$. Therefore, by our assumption, this collection will have non-empty intersection. Therefore, X is complete by Cantor's intersection theorem.

5.7 Completeness and Sequences of Bounded Variation

In this section, we give a characterisation of compleness in terms of sequences of bounded variations.

Definition 5.7.1 Let (X, d) be a metric space. A sequence $\langle x_n\rangle$ in X is said to be of bounded variation if the series $\sum_{n=1}^{\infty} d(x_n, x_{n+1})$ is convergent. A sequence $\langle x_n\rangle$ in X is said to satisfy the condition (C) with respect to a positive real number M and a real number r such that $0 < r < 1$ if

$$d(x_n, x_{n+1}) \leq Mr^n \text{ for all } n \in \mathbb{Z}_+. \qquad \text{(C)}$$

Lemma 5.7.2 1. If a sequence $\langle x_n \rangle$ in a metric space (X, d) satisfies the condition (C) with respect to a positive real number M and a real number r such that $0 < r < 1$, then $\langle x_n \rangle$ is of bounded variation.

2. Every sequence of bounded variation is a Cauchy sequence.

Proof. 1. By condition (C), for each $k \in \mathbb{Z}_+$,

$$\sum_{n=1}^{k} d(x_n, x_{n+1}) \leq M \sum_{n=1}^{k} r^n.$$

Since $\sum_{n=1}^{\infty} r^n$ is convergent for $0 < r < 1$, the series $\sum_{n=1}^{\infty} d(x_n, x_{n+1})$ is convergent, and $\langle x_n \rangle$ is of bounded variation.

2. Let $\langle x_n \rangle$ be a sequence of bounded variation in a metric space (X, d). Let $\epsilon > 0$ be a real number. Since $\langle x_n \rangle$ is of bounded variation, $\sum_{n=1}^{\infty} d(x_n, x_{n+1})$ is convergent. Then there exists a $p \in \mathbb{Z}_+$ such that

$$\sum_{n=p}^{\infty} d(x_n, x_{n+1}) < \epsilon.$$

Now, for $n \geq m \geq p$, the triangle inequality gives the inequality

$$d(x_n, x_m) \leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) < \epsilon.$$

This completes the proof.

Theorem 5.7.3 In a metric space (X, d), the following statements are equivalent.

1. X is complete.
2. Every sequence of bounded variation in X converges in X.
3. Every sequence in X that satisfies the condition (C) with respect to a positive real number M and a real number r such that $0 < r < 1$ converges.
4. For some positive real number M_0 and for some real number r_0 such that $0 < r_0 < 1$, every sequence in X satisfying the condition (C) with respect to M_0 and r_0 converges in X.

Proof. By Lemma 5.7.2, the proofs of (1) implies (2), and (2) implies (3) are obvious. Also the implication: (3) implies (4) is obvious. Thus, we prove that (4) implies (1). Let $\langle x_n \rangle$ be a Cauchy sequence in X. For each $k \in \mathbb{Z}_+$, there exists an $l_k \in \mathbb{Z}_+$ such that

$$d(x_n, x_m) < M_0 r_0^k \text{ for all } n, m \geq l_k.$$

We define a sequence $\langle n_k \rangle_{k=l}^{\infty}$ inductively as follows:

$$n_1 = l_1 \quad \text{and} \quad n_{k+1} = l_{k+1} \vee (n_k + 1).$$

Then the subsequence $\langle x_{n_k} \rangle_{k=1}^{\infty}$ of $\langle x_n \rangle$ satisfies the condition

$$d(x_{n_k}, x_{n_{k+1}}) < M_0 r_0^k \text{ for all } k \in \mathbb{Z}_+.$$

Therefore, $\langle x_n \rangle$ is convergent.

5.8 Completeness and Absolutely Closed Metric Spaces

This section too obtains an equivalent condition for completeness. Recall the definition of embedding of a metric space into another metric space from Section 5.4.

Definition 5.8.1 A metric space is said to be absolutely closed if its embedding is a closed subset of any metric space in which it can be embedded.

Theorem 5.8.2 A metric space is complete if and only if it is absolutely closed.

Proof. Let (X, d) be a complete metric space, and (X^*, d^*) be a metric space in which (X, d) is embedded by an isometry f. Let y be a limit point of $f(X)$. Then we can extract a sequence $\langle y_n \rangle$ in $f(X)$ which converges to y in X^*. This means that there is a sequence $\langle x_n \rangle$ in X such that $f(x_n) = y_n$ for each $n \in \mathbb{Z}_+$. Since f is an isometry from X onto $f(X)$

$$d(x_n, x_m) = d^*(f(x_n), f(x_m)) = d^*(y_n, y_m).$$

Since $\langle y_n \rangle$ converges to y in X^*, $\langle y_n \rangle$ is a Cauchy sequence, and therefore $\langle x_n \rangle$ is a Cauchy sequence in X. Since X is complete, $\langle x_n \rangle$ converges to a point x in X. Since f is an isometry, it is continuous, and so $\langle f(x_n) \rangle$ converges to $f(x) \in f(X)$. That is, $\langle y_n \rangle$ converges to $f(x)$. Therefore, $f(x) = y$, and $y \in f(X)$. Thus, $f(X)$ is a closed set in X^*, and (X, d) is absolutely closed. Conversely, suppose that (X, d) is an absolutely closed metric space. By Theorem 5.4.1, there is a complete metric space (X^*, d^*) and an isometry f from X onto a dense subset of (X^*, d^*). That is, X is embedded in X^* by the isometry f. By our assumption, $f(X)$ is a closed set in X^*, and hence it is a complete metric space. Since f is an isometry, X is also complete.

5.9 Completeness and Totally Bounded Sets

This section contains still another characterisation of completeness in terms of totally bounded sets.

Theorem 5.9.1 A subset A of a metric space (X, d) is totally bounded if and only if each sequence in A has a Cauchy subsequence.

Proof. First assume that A is a totally bounded subset of X. Let $\langle x_n \rangle$ be a sequence in A. By Lemma 2.17.4, there exists a partition of A into a finite number of subsets each having diameter less than one. Since these sets are finite in number, there is at least one set B_1 in this partition such that x_n is in B_1 for inifnitely many values of n. Thus, there is a subsequence $\langle x_n^1 \rangle$ of $\langle x_n \rangle$ which is in B_1. Since B_1 is contained in A, it is totally bounded. Therefore, there exists a partition of B_1 such that the diameter of each set in the partition is less than $\frac{1}{2}$. Similarly, as above, there exists a set B_2 in partition of B_1 such that x_n^1 is in B_2 for infinitely many values of n, that is, there exists a subsequence $\langle x_n^2 \rangle$ of $\langle x_n^1 \rangle$ which is in B_2. By induction, for each $n \in \mathbb{Z}_+$, there exists a set $B_n \subseteq B_{n-1}$ with diam $(B_n) < \frac{1}{n}$ such that the subsequence $\langle x_k^n \rangle$ of $\langle x_k^{n-1} \rangle$ lies in B_n.

Consider now the diagonal sequence $\langle x_n^n \rangle_{n=1}^{\infty}$ in the infinite array

$$x_1^1, x_2^1, x_3^1, \ldots, x_k^1, \ldots$$
$$x_1^2, x_2^2, x_3^2, \ldots, x_k^2, \ldots$$
$$\vdots$$
$$x_1^n, x_2^n, x_3^n, \ldots, x_k^n, \ldots$$
$$\vdots$$

The sequence $\langle x_n^n \rangle$ is a subsequence of x_n with the property that for each $m \in \mathbb{Z}_+$, the tail $\langle x_m^m, x_{m+1}^{m+1}, \ldots \rangle$ of $\langle x_n^n \rangle$ is a subsequence of $\langle x_k^m \rangle$, and therefore lies wholly in B_m. Hence, $\langle x_n^n \rangle$ is a Cauchy sequence.

Conversely, suppose on the contrary that A is not totally bounded. Then there exists a real number $\epsilon > 0$ such that there is no finite ϵ-net for A. Let x_1 be a point in A. Then A is not contained in $S(x_1, \epsilon)$, for otherwise the set $\{x_1\}$ is a finite ϵ-net for A. Therefore, there exists a point x_2 in A such that $d(x_1, x_2) \geq \epsilon$. Again A is not contained in $S(x_1, \epsilon) \cup S(x_2, \epsilon)$, for otherwise the set $\{x_1, x_2\}$ is a finite ϵ-net for A . Therefore, there exists a point x_3 in A such that $d(x_1, x_3) \geq \epsilon$ and $d(x_2, x_3) \geq \epsilon$. By induction, now there exists a sequence $\langle x_n \rangle$ in A such that $d(x_i, x_j) \geq \epsilon$ for $i \neq j$. Thus, $\langle x_n \rangle$ has no Cauchy subsequence, and the proof is complete.

Theorem 5.9.2 A metric space is complete if and only if every infinite totally bounded subset has a limit point.

Proof. Let (X, d) be a complete metric space, and A be an infinite totally bounded subset of X. Since A is infinite, a sequence $\langle x_n \rangle$ of distinct points can be extracted from A. By Theorem 5.9.1, $\langle x_n \rangle$ has a Cauchy subsequence $\langle x_{n_k} \rangle_{k=1}^{\infty}$. Since X is complete, $\langle x_{n_k} \rangle$ converges to a point x in X. Therefore, x is a limit point of A. Conversely, assume that every infinite totally bounded subset in a metric space (X, d) has a limit point. Let $\langle x_n \rangle$ be a Cauchy sequence in X. Let A be the set of points in $\langle x_n \rangle$. If A is finite, then some point of $\langle x_n \rangle$ is repeated infinite number of times in the sequence $\langle x_n \rangle$, that is, there exists a $k \in \mathbb{Z}_+$ such that $x_n = x_k$ for infinitely many values of n. But then there exists a constant subsequence of $\langle x_n \rangle$ consisting of x_k's which, of course, converges to x_k. Thus, $\langle x_n \rangle$ converges to x_k in this case. So assume that A is infinite. Also A is totally bounded since for each real number $\epsilon > 0$, there exists a $p \in \mathbb{Z}_+$ such that $d(x_n, x_m) < \epsilon$ for all $n, m \geq p$, and so $A \subseteq \bigcup_{n=1}^{p} S(x_n, \epsilon)$, that is, the set $\{x_1, x_2, \ldots, x_p\}$ is an ϵ-net. By our assumption, A has a limit point $x \in X$. Clearly, $\langle x_n \rangle$ converges to x. Thus, X is complete.

5.10 Completeness and Mappings

Theorem 5.10.1 Completeness is preserved under isometries.

Proof. Let (X, d_1) be a complete metric space, and (Y, d_2) be a metric space. Let $f : X \to Y$ be an isometry. We show that $f(X)$ is a complete subspace of Y. Let $\langle y_n \rangle$ be a Cauchy sequence in $f(X)$. For each n, there exists $x_n \in X$ such that $f(x_n) = y_n$.

Since f is an isometry,

$$d_2(y_n, y_m) = d_2(f(x_n), f(x_m)) = d_1(x_n, x_m).$$

From the above inequality, it follows that $\langle x_n \rangle$ is a Cauchy sequence. Since X is complete, $\langle x_n \rangle$ converges to a point x in X. Now the equality

$$d_1(x_n, x) = d_2(f(x_n), f(x)) = d_2(y_n, f(x))$$

shows that $\langle y_n \rangle$ converges to $f(x) \in f(X)$. Thus, $f(X)$ is complete.

The continuous mappings need not preserve completeness, not even homeomorphisms. The mapping $f : \mathbb{R} \to]-1, 1[$ defined by $f(x) = \frac{x}{(1+|x|)}$ is a homeomorphism. Also, $\mathbb{R}$ is complete while $]-1, 1[$ is not. The uniformly continuous mappings also need not preserve completeness. Let d' be he discrete metric on $\mathbb{R}$, and d be the usual metric on $\mathbb{R}$. Then any mapping from $(\mathbb{R}, d')$ into $(\mathbb{R}, d)$ is uniformly continuous. In particular, if we take the mapping $f : (\mathbb{R}, d') \to (\mathbb{R}, d)$ defined by $f(x) = \frac{x}{(1+|x|)}$, then $f(\mathbb{R}) =]-1, 1[$, which is not complete.

Example 5.10.2 We have already constructed a homeomorphism between $\mathbb{R}$ and an open interval $]a, b[$ in $\mathbb{R}$. So let $f : \mathbb{R} \to]a, b[$ be a homeomorphism. We use f to define

$$\sigma(f(x), f(y)) = |x - y| \quad \text{for all} \quad x, y \text{ in } \mathbb{R}.$$

It is easily seen that σ is a new metric on $[a, b[$, and with respect to this metric on $]a, b[$, $f : \mathbb{R} \to]a, b[$ is an isometry. Since $\mathbb{R}$ is complete, $]a, b, [$ is complete with respect to σ. We show that the usual metric d and the metric σ on $]a, b[$ are equivalent. Let $\langle x_n \rangle$ be a sequence in $(]a, b[, d)$ converging to $x \in]a, b[$. Since f is a homeomorphism, the sequence $\langle f^{-1}(x_n) \rangle$ in $\mathbb{R}$ converges to $f^{-1}(x)$. Again since f is an isometry from $\mathbb{R}$ onto $(]a, b[, \sigma)$, the sequence $\langle f(f^{-1}(x_n)) \rangle$ converges in $(]a, b[, \sigma)$ to $f(f^{-1}(x))$. This means that the sequence $\langle x_n \rangle$ converges in $(]a, b[, \sigma)$ to x. Conversely, let $\langle y_n \rangle$ be a convergent sequence in $(]a, b[, \sigma)$ converging to $y \in]a, b[$. Since f is an isometry, the sequence $\langle f^{-1}(y_n) \rangle$ in $\mathbb{R}$ will converge to $f^{-1}(y) \in \mathbb{R}$. As $f : \mathbb{R} \to (]a, b[, d)$ is a homeomorphism, the sequence $\langle f(f^{-1}(y_n)) \rangle$ converges in $(]a, b[, d)$ to $f(f^{-1}(y))$, that is, $\langle y_n \rangle$ converges in $(]a, b[, d)$ to y . Thus, by Theorem 2.6.7, d and σ on $]a, b[$ are equivalent. Now with respect to d, the interval $[a, b[$ is not complete, and with respect to σ the interval $]a, b[$ is complete. This example shows that completeness is not a topological property.

The points made in the above example can be stated more generally. Let X be any set, and (Y, ρ) be a metric space and $\phi : X \to Y$ be a bijective mapping. Then we can define a mapping $\sigma : X \times X \to \mathbb{R}$ by setting

$$\sigma(x, y) = \rho(\phi(x), \phi(y)) \quad \text{for all} \quad x, y \text{ in } X.$$

Then σ is indeed a metric on X. Also $\phi : (X, \sigma) \to (Y, \rho)$ is an isometry, and (X, σ) inherits all the properties of (Y, ρ). If d is another metric on X such that $\phi : (X, d) \to (Y, \rho)$ is a homeomorphism, then d and σ are equivalent.

The next theorem investigates the problem of extension of uniformly continuous mappings.

Theorem 5.10.3 Let (X, d) be a metric space, and A be a dense subset of X. Let (Y, ρ) be a complete metric space, and $f : A \to Y$ be a uniformly continuous mapping. Then f can be extended uniquely to uniformly continuous mapping $\bar{f} : X \to Y$.

Proof. Let x be any point in X. Since $\bar{A} = X$, there exists a sequence $\langle x_n \rangle$ in A such that $\langle x_n \rangle$ converges to x. Then $\langle x_n \rangle$ is a Cauchy sequence in A. Since f is uniformly continuous, $\langle f(x_n) \rangle$ is a Cauchy sequence in Y. By the completeness of Y, the sequence $\langle f(x_n) \rangle$ converges to a point $y \in Y$. We define $\bar{f} : X \to Y$ by $\bar{f}(x) = y$. We show that $\bar{f}$ is well-defined. For this we need to show that if $\langle x'_n \rangle$ is another sequence in A converging to x, then the sequence $\langle f(x'_n) \rangle$ converges in Y to y. Let $\epsilon > 0$ be a real number. Since f is uniformly continuous, there exists a real number $\delta > 0$ such that for all x, x' in A

$$d(x, x') < \delta \quad \text{implies that} \quad \rho(f(x), f(x')) < \frac{\epsilon}{2}.$$

Since both $\langle x_n \rangle$ and $\langle x'_n \rangle$ converges in X to x, and

$$d(x_n, x'_n) \leq d(x_n, x) + d(x, x'_n),$$

then there exists a $p \in \mathbb{Z}_+$ such that $d(x_n, x'_n) < \delta$ for all $n \geq p$. Then $\rho(f(x_n), f(x'_n)) < \frac{\epsilon}{2}$ for all $n \geq p$. Again since $\langle f(x_n) \rangle$ converges to y, from the inequality

$$\rho(f(x'_n), y) \leq \rho(f(x'_n), f(x_n)) + \rho(f(x_n), y)$$

it follows that there is a $p_0 \in \mathbb{Z}_+$ such that $\rho(f(x'_n), y) < \epsilon$ for all $n \geq p_0$. Thus, $\bar{f}$ is well-defined. Now we show that $\bar{f}(x) = f(x)$ for all x in A so that $\bar{f}$ is an extension of f. For any point x in A, the constant sequence $\langle x, x, \dots \rangle$ is in A, and it converges to x. Its image sequence under f is $\langle f(x), f(x), \dots \rangle$ which is also constant sequence, and thus converges to $f(x)$. Therefore, $\bar{f}(x) = f(x)$ for all x in A. Now we show that $\bar{f}$ is uniformly continuous. Since f is uniformly continuous on A, there exists a real number $\delta > 0$ such that for all x, x' in A

$$d(x, x') < \delta \quad \text{implies that } \rho(f(x), f(x')) < \frac{\epsilon}{3}.$$

Now let x and x' be any points in X. Then there exist sequences $\langle x_n \rangle$ and $\langle x'_n \rangle$ in A converging in X to x and x', respectively. Then there exists a $p_1 \in \mathbb{Z}_+$ such that

$$d(x_n, x) < \frac{\delta}{3} \quad \text{and } d(x'_n, x') < \frac{\delta}{3} \quad \text{for all } n \geq p_1.$$

Then by the inequality

$$d(x_n, x'_n) \leq d(x_n, x) + d(x, x') + d(x', x'_n)$$

it follows that $d(x_n, x'_n) < \delta$ for all $n \geq p_1$ whenever $d(x, x') < \frac{\delta}{3}$. Therefore,

$$\rho(f(x_n), f(x'_n)) < \frac{\epsilon}{3} \quad \text{for all } n \geq p_1 \text{ whenever } d(x, x') < \frac{\delta}{3}.$$

Since $\langle f(x_n) \rangle$ and $\langle f(x'_n) \rangle$ converge to $\bar{f}(x)$ and $\bar{f}(x')$, respectively, there exists a $p_2 \in \mathbb{Z}_+$ such that

$$\rho(f(x_n), \bar{f}(x)) < \frac{\epsilon}{3} \quad \text{and} \quad \rho(f(x'_n), \bar{f}(x')) < \frac{\epsilon}{3} \quad \text{for all } n \geq p_2.$$

Then

$$\begin{aligned} \rho(\bar{f}(x), \bar{f}(x')) &\leq \rho(\bar{f}(x), f(x_n)) + \rho(f(x_n), f(x'_n)) + \rho(f(x'_n), \bar{f}(x')) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \quad \text{for all } n \geq \max\{p_1, p_2\} \text{ whenever } d(x, x') < \frac{\delta}{3}. \end{aligned}$$

Thus, $\bar{f}$ is uniformly continuous on X. It remains to show that $\bar{f}$ is unique. Let $\bar{g}$ be any other uniformly continuous extension of f. Let x be any point in X. Since $\bar{A} = X$, there exists a sequence $\langle x_n \rangle$ in A such that $\langle x_n \rangle$ converges to x in X. Since $\bar{f}$ and $\bar{g}$ are uniformly continuous, $\langle \bar{f}(x_n) \rangle$ converges to $\bar{f}(x)$, and $\langle \bar{g}(x_n) \rangle$ converges to $\bar{g}(x)$. Then there exists a $p_0 \in \mathbb{Z}_+$ such that

$$\rho(f(x_n), \bar{f}(x)) = \rho(\bar{f}(x_n), \bar{f}(x)) < \frac{\epsilon}{2} \quad \text{for all } n \geq p_0$$

and

$$\rho(f(x_n), \bar{g}(x)) = \rho(\bar{g}(x_n), \bar{g}(x)) < \frac{\epsilon}{2} \quad \text{for all } n \geq p_0.$$

Then

$$\rho(\bar{f}(x), \bar{g}(x)) \leq \rho(\bar{f}(x), f(x_{p_0})) + \rho(f(x_{p_0}), \bar{g}(x)) < \epsilon.$$

Since ϵ is arbitrary, $\bar{f}(x) = \bar{g}(x)$.

When A is not dense in the above theorem, we have the following result.

Theorem 5.10.4 Let f be a uniformly continuous mapping of a subset A of a metric space (X, d) into complete metric space (Y, ρ). Then the subset of X to which f extends continuously coincides with $\bar{A}$ in X.

Proof. Let $\epsilon > 0$ be a real number. Since f is uniformly continuous on A, there exists a real number $\delta > 0$ such that $d(x, x') < \delta$ implies that $\rho(f(x), f(x')) < \epsilon$ for all x and x' in A. Let x_0 be a point in $\bar{A}$. Then the open sphere $S(x_0, \frac{\delta}{2})$ intersects with A. Then for x and x' in $S(x_0, \frac{\delta}{2}) \cap A$, it follows that $\rho(f(x), f(x')) < \epsilon$. Thus, the diameter $\operatorname{diam}(f(S(x_0, \frac{\delta}{2}) \cap A)) \leq \epsilon$. So it follows that $\omega(f; x_0) = 0$, and hence $\lim\limits_{\substack{x \to x_0 \\ x \in A}} f(x)$ exists.

EXERCISE

1. Let m and n be positive integers, and $F_{m,n}$ be the subset of $\mathbb{R}^2$ given by

$$F_{m,n} = \{(x, y) \in \mathbb{R}^2 : |x| > m \quad \text{and } |y| > n\}.$$

Show that the collection $\{F_{m,n}\}$ has the finite intersection property, and further show that $\bigcap_{m,n} F_{m,n}$ is empty.

5.11 Spaces of Continuous Mappings

Recall the metric space $\mathcal{C}[a, b]$ of real valued continuous mappings defined on the closed intervals $[a, b]$ with sup metric $d(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$. This metric space is important from one more point of view that it is a vector space (or a linear space) over the field of real numbers with respect to pointwise addition and scalar multiplication

$$(f + g)(x) = f(x) + g(x),$$

$$(\alpha f)(x) = \alpha f(x)$$

where $f, g \in \mathcal{C}[a, b]$ and α is a scalar in $\mathbb{R}$, due to the following properties:

1. $f + g \in \mathcal{C}[a, b]$ for all $f, g \in \mathcal{C}[a, b]$.
2. $(f + g) + h = f + (g + h)$ for all f, g and h in $\mathcal{C}[a, b]$.
3. The zero function given by $0(x) = 0$ for all $x \in [a, b]$ is in $\mathcal{C}[a, b]$, and it has the property that
$$0 + f = f + 0 = f \quad \text{for all } f \text{ in } \mathcal{C}[a, b].$$
The function 0 is called the additive identity of $\mathcal{C}[a, b]$.
4. For each $f \in \mathcal{C}[a, b]$, $(-1)f \in \mathcal{C}[a, b]$, we write $(-1)f = -f$, and it is such that
$$f + (-f) = 0 = (-f) + f.$$
$-f$ is called is called the additive inverse of f.
5. $f + g = g + f$ for all $f, g \in \mathcal{C}[a, b]$.
6. If $f \in \mathcal{C}[a, b]$ and $\alpha \in \mathbb{R}$, then $\alpha f \in \mathcal{C}[a, b]$.
7. $(\alpha + \beta)f = \alpha f + \beta f$ for all $\alpha, \beta \in \mathbb{R}$ and $f \in \mathcal{C}[a, b]$.
8. $\alpha(f + g) = \alpha f + \alpha g$ for all $\alpha \in \mathbb{R}$ and $f, g \in \mathcal{C}[a, b]$.
9. $(\alpha\beta)f = \alpha(\beta f)$ for all $\alpha, \beta \in \mathbb{R}$ and $f \in \mathcal{C}[a, b]$.
10. $1f = f$, where 1 is the multiplicative identity of $\mathbb{R}$.

We also define a mapping $\| \| : \mathcal{C}[a, b] \to \mathbb{R}$, by setting
$$\|f\| = \sup\{|f(x)| : x \in [a, b]\}$$
where we agree to write $\| \|(f) = \|f\|$. This mapping is actually a norm since it satisfies the following properties.

1. $\|f\| \geq 0$, and $\|f\| = 0$ if and only if $f = 0$.
2. $\|f + g\| \leq \|f\| + \|g\|$.
3. $\|\alpha f\| = |\alpha| \, \|f\|$.

Further, if we agree to write $f + (-g) = f - g$, then the metric d is induced by this norm since
$$d(f, g) = \|f - g\|.$$
Therefore, $\mathcal{C}[a, b]$ is a normed vector space which is also a metric space with its metric induced by its norm.

Definition 5.11.1 A Banach space is a normed vector space which is complete as a metric space with its metric induced by its norm.

Now since $\mathcal{C}[a, b]$ is complete with its metric d induced by its norm, it is a Banach space.

Since any closed subspace of a complete metric space is complete, any closed vector subspace of a Banach space is complete, and hence itself is a Banach space with respect to the same algebraic operations and the norm. Thus, any closed vector subspace of the Banach space $\mathcal{C}[a, b]$ is a Banach space.

The property which has been utilised in the above discussion of $\mathcal{C}[a,b]$ to define its norm is the fact that a continuous real valued function defined on a closed interval $[a,b]$ is bounded, and therefore the norm $\|f\|$ and hence the metric is well-defined. If we replace the closed interval $[a,b]$ by any metric space (X,d), then $\|f\|$ may not be well-defined, since $\sup\{\|f(x)\| : x \in X\}$ may not exist. However, if we consider continuous mappings which are bounded also, then the above problem is solved, and we still get a Banach space. This Banach space is denoted by $\mathcal{C}[X,\mathbb{R}]$. The following discussion is devoted to establish this fact more explicitely.

Suppose that X is any non-empty set. Let L be the set of all mappings $f : X \to \mathbb{R}$. Then L is a vector space over the field of real numbers $\mathbb{R}$ called a real vector space, with respect to pointwise addition and scalar multiplication: $(f+g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$. Now let B be the subset of L consisting of all bounded mappings. Then B is a vector subspace of L with respect to the same algebraic operations as are available in L, and therefore B is a vector space in its own right. Moreover, we can define norm of B by $\|f\| = \sup\{|f(x)| : x \in X\}$. Then B is a normed vector space, and it is a metric space with its metric d induced by its norm, that is, $d(f,g) = \|f-g\|$. Also the metric space (B,d) is complete by Example 5.1.15. Thus, B is a Banach space of all bounded real valued mappings which is also a subspace of L.

Now let X be a metric space, and $\mathcal{C}_b(X,\mathbb{R})$ be the subset of B consisting of continuous and bounded mappings. Since the constant mappings are continuous, $\mathcal{C}_b(X,\mathbb{R})$ is non-empty. With algebraic operations of addition and scalar multiplication as defined in B, the following lemma is straight forward.

Lemma 5.11.2 Let f and g be any elements in $\mathcal{C}_b(X,\mathbb{R})$ and α be a scalar in $\mathbb{R}$. Then $f+g \in \mathcal{C}_b(X,\mathbb{R})$, and $\alpha f \in \mathcal{C}_b(X,\mathbb{R})$.

By the above lemma, it follows that $\mathcal{C}_b(X,\mathbb{R})$ is a vector subspace of B, and also as a metric space, $\mathcal{C}_b(X,\mathbb{R})$ is a subspace of the metric space B. Now we show that $\mathcal{C}_b(X,\mathbb{R})$ is a closed set in the metric space B.

Lemma 5.11.3 $\mathcal{C}_b(X,\mathbb{R})$ is a closed subset of the metric space B.

Proof. Assume that the metric on X is d'. Let $x_0 \in X$. Let f be an element of B which is in the closure of $\mathcal{C}_b(X,\mathbb{R})$. We show that f is continuous at x_0. Let $\epsilon > 0$ be a real number. Then there exists an element g in $\mathcal{C}_b(X,\mathbb{R})$ such that $d(f,g) < \frac{\epsilon}{3}$. Since g is continuous, there exists a real number $\delta > 0$ such that for all $x \in X$

$$d'(x,x_0) < \delta \quad \text{implies that } |g(x) - g(x_0)| < \frac{\epsilon}{3}.$$

Then $d'(x,x_0) < \delta$ implies that

$$\begin{aligned} |f(x) - f(x_0)| &\le |f(x) - g(x)| + |g(x) - g(x_0)| + |g(x_0) - f(x_0)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Therefore, f is continuous at x_0. Since x_0 is arbitrary, f is continuous.

Since B is complete, by the above lemma, it follows that $\mathcal{C}_b(X,\mathbb{R})$ is complete, and so it is a Banach space. Thus, we established the following.

Theorem 5.11.4 $\mathcal{C}_b(X,\mathbb{R})$ is a Banach space.

A moment of thought will convince the reader that in the entire discussion carried above if we replace $\mathbb{R}$ by the complex plane $\mathbb{C}$, that is, consider the complex valued mappings in

place of real valued mappings, we obtain the corresponding vector spaces. Thus, if we let $\mathcal{C}_b(X, \mathbb{C})$ to denote the set of complex valued continuous and bounded mappings defined on X, with pointwise addition, scalar multiplication and sup norm: $\|f\| = \sup\{|f(x)| : x \in X\}$ we consider that the following statement is proved.

Theorem 5.11.5 $\mathcal{C}_b(X, \mathbb{C})$ is a Banach space.

5.12 Completeness and Filters

In this section, we obtain a characterisation of completeness in terms of filters.

Definition 5.12.1 A filter on a set X is a set $\mathcal{B}$ of subsets of X which has the following properties.

1. Every subset of X which contains a set of $\mathcal{B}$ belongs to $\mathcal{B}$.
2. Every finite intersection of sets in $\mathcal{B}$ belongs to $\mathcal{B}$.
3. The empty set is not in $\mathcal{B}$.

It follows from the statements (2) and (3) in the above definition that every finite intersection of sets in $\mathcal{B}$ is non-empty. Also a filter on X always contains X; it is the intersection of the empty collection of sets in $\mathcal{B}$.

Examples 5.12.2

1. A set $\mathcal{B}$ of subsets of a set X such that each $B \in \mathcal{B}$ contains a given set A is a filter. Also, if X is non-empty, then the set $\{X\}$ is a filter on X.
2. Let (X, d) be a metric space, and $x \in X$. The set of all neighbourhoods of x is a filter called the neighbourhood filter of x.
3. If X is an infinite set, and $\mathcal{B}$ is the set of complements of finite subsets of X, then $\mathcal{B}$ is a filter. If $X = \mathbb{N}_0$, then $\mathcal{B}$ is known as Fréchet filter.

Definition 5.12.3 Given two filters $\mathcal{B}$ and $\mathcal{B}'$ on a set X, $\mathcal{B}'$ is said to be finer than $\mathcal{B}$, or $\mathcal{B}$ is coarser than $\mathcal{B}'$ if $\mathcal{B} \subseteq \mathcal{B}'$.

Definition 5.12.4 Let $\mathcal{B}$ be a set of subsets of a set X. Then $\mathcal{B}$ is called a filter base on X if the following two conditions hold.

1. The intersection of two sets of $\mathcal{B}$ contains a set of $\mathcal{B}$.
2. $\mathcal{B}$ is non empty and the empty subset of X is not in $\mathcal{B}$.

Theorem 5.12.5 A set $\mathcal{B}$ of subsets of a set X is a filter base on X if and only if the set $\mathcal{B}' = \{A \subseteq X : \text{for some } B \in \mathcal{B}, B \subseteq A\}$ is a filter.

Proof. Assume that $\mathcal{B}$ is a filter base on X. Let U be a subset of X such that there is a set B in $\mathcal{B}'$ which is contained in U. Since B is in $\mathcal{B}'$, there exists a set C in $\mathcal{B}$ which is contained in B. Then C is contained in U. Therefore, U is in $\mathcal{B}'$. Let $U_1, U_2, \ldots, U_n$ be sets in $\mathcal{B}'$. For each $i = 1, 2, \ldots, n$, there exists a set B_i in $\mathcal{B}$ such that B_i is contained in U_i.

Since $\mathcal{B}$ is a filter base, there exists a set B in $\mathcal{B}$ such that $B \subseteq \bigcap_{i=1}^{n} B_i$. Thus, $B \subseteq \bigcap_{i=1}^{n} U_i$. So that $\bigcap_{i=1}^{n} U_i$ is in $\mathcal{B}'$. Since $\mathcal{B}$ does not contain the empty set, $\mathcal{B}'$ does not contain the empty set. Thus, $\mathcal{B}'$ is a filter. Conversely, assume that $\mathcal{B}'$ is a filter. Let U_1 and U_2 be two element in $\mathcal{B}$. Then U_1 and U_2 are in $\mathcal{B}'$. Since $\mathcal{B}'$ is a filter, $U_1 \cap U_2$ is in $\mathcal{B}'$. Therefore, there exists a set $B \in \mathcal{B}$ such that $B \subseteq U_1 \cap U_2$. If $\mathcal{B}$ contains the empty set, then $\mathcal{B}'$ contains the empty set, a contradiction to the fact that $\mathcal{B}'$ is a filter. If $\mathcal{B}$ is empty, then this would entail that $\mathcal{B}'$ is empty, a contradiction since a filter on X always contain the set X.

The filter $\mathcal{B}'$ in the above theorem is said to be generated by the filter base $\mathcal{B}$.

Definition 5.12.6 Let (X, d) be a metric space, and $\mathcal{B}$ a filter on X. A point $x \in X$ is said to be a limit of $\mathcal{B}$, written $\lim \mathcal{B} = x$ if $\mathcal{B}$ is finer than the neighbourhood filter $\mathcal{N}(x)$ of x. When $\lim \mathcal{B} = x$, we sometimes say that $\mathcal{B}$ converges to x.

Definition 5.12.7 Let (X, d) be a metric space. A point x in X is said to be a limit of a filter base $\mathcal{B}$ on X, written $\lim \mathcal{B} = x$, or $\mathcal{B}$ converges to x, if the filter whose filter base is $\mathcal{B}$ converges to x.

Definition 5.12.8 A filter base $\mathcal{B}$ on a metric space X is Cauchy or satisfies the Cauchy criterion if $\inf\{\text{diam}(B) : B \in \mathcal{B}\} = 0$.

The necessary preliminaries are now ready for the following characterisation of completeness we were looking for.

Theorem 5.12.9 A metric space (X, d) is complete if and only if every Cauchy filter base on X is convergent in X.

Proof. Let (X, d) be a metric space such that every Cauchy filter base on X is convergent in X. Let $\langle x_n \rangle$ be a Cauchy sequence in X. For each k, let $A_k = \{x_n; n \geq k\}$. Then the set $\mathcal{B} = \{A_k : k \in \mathbb{Z}_+\}$ is a filter base on X. Let $\epsilon > 0$ be a real number. Then there exists a $k \geq \mathbb{Z}_+$ such that $d(x_n, x_m) < \frac{\epsilon}{2}$ for all $n, m \geq k$. This means that $\text{diam}(A_k) < \epsilon$. Since ϵ is arbitrary, it follows that $\inf\{\text{diam}(A_k) : k \in \mathbb{Z}_+\} = 0$. Thus, $\mathcal{B}$ is a Cauchy filter base. By our assumption, there exists a point x in X such that $\mathcal{B}$ converges to x. Now the set $\mathcal{B}' = \{A \subseteq X :$ there is some $A_k \in \mathcal{B}$ such that $A_k \subseteq A\}$ is the filter generated by $\mathcal{B}$. Then $\mathcal{B}'$ converges to x. This means that $\mathcal{B}'$ is finer than the neighbourhood filter $\mathcal{N}(x)$ of x. Since the open sphere $S(x, \epsilon)$ is in $\mathcal{N}(x)$, it is in $\mathcal{B}'$. Then, by the definition of $\mathcal{B}'$, there exists an A_k in $\mathcal{B}$ such that $A_k \subseteq S(x, \epsilon)$. But this means that $x_n \in S(x, \epsilon)$ for all $n \geq k$. Therefore, $\langle x_n \rangle$ converges to x in X, and X is complete. Conversely, suppose that (X, d) is a complete metric space, and $\mathcal{B}$ is a Cauchy filter base on X. Then for each $n \in \mathbb{Z}_+$, there exists a set A_n in $\mathcal{B}$ such that $\text{diam}(A_n) < \frac{1}{n}$. Choose, for each n, a point x_n from A_n. Then $d(x_n, x_m) \leq \frac{1}{n} + \frac{1}{m}$ since A_n and A_m in $\mathcal{B}$ must intersect as $\mathcal{B}$ is a filter base. This shows that $\langle x_n \rangle$ is a Cauchy sequence, and, by completeness of X, there is a point x in X to which $\langle x_n \rangle$ converges. Then for any real number $\epsilon > 0$, there is an $n_0 \in \mathbb{Z}_+$ such that $d(x_n, x) < \frac{\epsilon}{2}$ for all $n \geq n_0$. If $n \in \mathbb{Z}_+$ is such that $n > \max\{n_0, \frac{2}{\epsilon}\}$, then $x_n \in S(x, \frac{\epsilon}{2})$ and also $\text{diam}\, A_n < \frac{\epsilon}{2}$. It now follows using the triangle inequality that $A_n \subseteq S(x, \epsilon)$. Thus, $\lim \mathcal{B} = x$, and the proof is complete.

5.13 G_δ-Set in Complete Metric Spaces

Recall the definition of a G_δ-set from Exercise 19 in Section 5.5. We have the following.

Theorem 5.13.1 Every G_δ-set in a complete metric space is homeomorphic to a complete metric space.

Proof. Let (X, d) be a complete metric space, and A be a G_δ-set in X. Then $A = \bigcap_{i=1}^{\infty} G_i$, where G_i's are open sets in X. Let $f_i : G_i \to \mathbb{R}$ be a mapping defined by

$$f_i(x) = \frac{1}{d(x, X - G_i)} \quad \text{for all } x \text{ in } G_i.$$

Let $\rho_i : G_i \times G_i \to \mathbb{R}$ be a mapping defined by

$$\rho_i(x, y) = \frac{|f_i(x) - f_i(y)|}{1 + |f_i(x) - f_i(y)|}.$$

The mapping ρ_i may not be a metric on G_i since it is possible that $\rho_i(x, y)$ is zero for $x \neq y$. The triangle inequality for ρ_i can be proved as in Example 2.1.15. The desired metric ρ on A can be defined as

$$\rho(x, y) = d(x, y) + \sum_{n=1}^{\infty} \frac{\rho_n(x, y)}{2^n} \quad \text{for all } x \text{ and } y \text{ in } A.$$

The series in the above equality converges uniformly in A, and ρ is a metric on A. To see that (A, ρ) is complete, let $\langle x_n \rangle$ be a Cauchy sequence in (A, ρ). Then $\langle x_n \rangle$ is Cauchy sequence in (X, d). Since (X, d) is complete, $\langle x_n \rangle$ converges to a point x in X. If $x \in A$, then it is easy to see that $\langle x_n \rangle$ converges in (A, ρ) to x. It follows that every Cauchy sequence in (A, ρ) converges, and also that d and ρ are equivalent on A. If x does not belong to A, then there exists an $n_0 \in \mathbb{Z}_+$ such that $x \in X - G_n$ for $n > n_0$. Select a term x_k of the sequence $\langle x_n \rangle$. Then $\rho_i(x_k, x_{k+j})$ tends to 1 since x_{k+j} tends x, so that $d(x_{k+j}, X - G_i)$ tends to zero. But then $\rho(x_k, x_{k+j})$ has a limit not less than $\sum_{n=n_0}^{\infty} 2^{-n}$, and $\langle x_n \rangle$ can not be a Cauchy sequence.

EXERCISE

1. Let X and Y be complete metric spaces, and $A \subseteq X$ and $B \subseteq Y$. Let $f : A \to B$ be a homeomorphism. Then there exists an extension $\hat{f}$ of f such that $\hat{f}$ is a homeomorphism from a set $\hat{A} \subseteq \bar{A}$ in X onto a set $\hat{B} \subseteq \bar{B}$ in Y, where $\hat{A}$ and $\hat{B}$ are G_δ-sets in X and Y, respectively.

Chapter 6

Compactness

In this chapter, we introduce a class of metric spaces known as the class of compact metric spaces. This class is a subclass of the class of complete metric spaces. The concept of compactness of metric spaces is an abstraction of a property possessed by certain sets of real numbers, and is expressed by Heine-Borel theorem which asserts the following: if X is a closed and bounded subset of the real line $\mathbb{R}$, then any family of open sets in $\mathbb{R}$ whose union contains X has a finite sub family whose union also contains X. If X is metric space in its own right, then the above theorem can be thought as saying that any class of open sets in X whose union is equal to X has a finite subclass whose union is also equal to X. The metric spaces in which the conclusion of this theorem holds are known as compact metric spaces. In such metric spaces the continuous mappings are well-behaved. For example, any real valued continuous mapping defined on a compact metric space is automatically bounded and uniformly continuous. In contrast, a real valued continuous mapping defined on metric space which is not compact may fail to be bounded and uniformly continuous. For instance, the mapping f from $]0,1[$ into $\mathbb{R}$ defined by$f(x) = \frac{1}{x}$ is neither bounded nor uniformly continuous. This chapter begins with the precise formulation of the notion of compactness of metric spaces. Then the behaviour of continuous mappings on compact metric spaces is examined. As usual there are several properties that are equivalent to the notion of compactness, and an excursion into these equivalences reveal that the notion of compactness is central to the project of modern analysis.

6.1 Compact Metric Spaces

Definition 6.1.1 Let (X, d) be a metric space. Let $\mathbb{G} = \{G_i\}$ be a family of open sets in X. Then $\mathbb{G}$ is called an open cover of X if each point of X is in some G_i in $\mathbb{G}$, that is, $\bigcup_{G_i \in \mathbb{G}} G_i = X$. A subcollection $\mathbb{G}'$ of $\mathbb{G}$ whose union is also X, that is, $\bigcup_{G_i \in \mathbb{G}'} G_i = X$ is called a subcover of X in $\mathbb{G}$. If $\mathbb{G}'$ is finite, it is called a finite subcover of X in $\mathbb{G}$.

Definition 6.1.2 A metric space (X, d) is called compact if every open cover of X has a finite subcover of X.

Definition 6.1.3 Let (X, d) be a metric space, and Y be a non-empty subset of X. Then Y is said to be compact in X if Y is compact as a subspace of X.

According to the definition of compactness the empty set is not compact as it does not

form a subspace of a metric space.

Examples 6.1.4

1. The real line $\mathbb{R}$ is not compact. For each $n \in \mathbb{Z}_+$, let $G_n =]-n, n[$, then the collection $\mathbb{G} = \{G_n : n \in \mathbb{Z}_+\}$ is an open cover of $\mathbb{R}$ which has no finite subcover of $\mathbb{R}$.

2. The open interval $]a, b[$ with usual metric is not compact. Let c be any point in $]a, b[$. For each $n \in \mathbb{Z}_+$, let $\varepsilon_n = \frac{\min\{c-a, b-c\}}{2^n}$. Now let $G_n =]a + \varepsilon_n, b - \varepsilon_n[$ for each $n \in \mathbb{Z}_+$. Then $G = \{G_n : n \in \mathbb{Z}_+\}$ is an open cover of $]a, b[$ which has no finite subcover of $]a, b[$.

3. Any finite metric space is compact as there are only finite number of open sets.

4. $\mathbb{R}^2$ is not compact. Consider the open cover of $\mathbb{R}^2$ consisting of open sets $G_n = \{(x, y) \in \mathbb{R}^2 : -n < x < n$, and y is arbitrary$\}$ for each $n \in \mathbb{Z}_+$. Extend the same reasoning to show that $\mathbb{R}^n$ and $\mathbb{C}^n$ are not compact.

5. Any infinite discrete metric space X is not compact. The collection $\mathbb{G} = \{\{x\} : x \in X\}$ is an open cover of X which has no finite subcover of X.

We note that the notion of a compact set$A$ in a metric space X is defined in terms of open sets in the subspace A. The following theorem shows that the open sets in X can be used instead.

Theorem 6.1.5 Let (X, d) be a metric space, and A be a non-empty subset of X. Then A is compact if and only if every covering of A by means of open sets of X contains a finite subcovering of A.

Proof. Let $\mathbb{G} = \{G_i\}_{i\in\Lambda}$ be a covering of A consisting of open sets in X, that is, $A \subseteq \bigcup_{i\in\Lambda} G_i$. For each $i \in \Lambda$, since G_i is open in X, $G_i \cap A$ is open in A. Then the collection $\mathbb{G}' = \{G_i \cap A\}_{i\in\Lambda}$ is an open cover of A. Since A is compact, there exists a finite subcover of A in $\mathbb{G}'$, say,

$$\{G_{i_1} \cap A, G_{i_2} \cap A, \ldots, G_{i_n} \cap A\}.$$

Then $A \subseteq \bigcup_{j=1}^{n} G_{i_j}$, and the set $\{G_{i_1}, G_{i_2}, \ldots, G_{i_n}\}$ is a finite subcover of A in G. Conversely, let $\mathbb{G} = \{G_i\}$ be an open cover of A consisting of open sets in A. For each $G_i \in \mathbb{G}$, there exists an open set G_i^* in X such that $G_i = G_i^* \cap A$. Then the collection of G_i^* obtained in this way forms a covering of A. By our assumption, this open cover of A has a finite subcover of A, say,

$$\left\{G_{j_1}^*, G_{j_2}^*, \ldots, G_{j_k}^*\right\}.$$

Then $A \subseteq \bigcup_{i=1}^{k} G_{j_i}^*$. Then $A = \bigcup_{i=1}^{k} \left(G_{j_i}^* \cap A\right) = \bigcup_{i=1}^{n} G_{j_i}$. Thus, $\{G_{j_1}, G_{j_2}, \ldots, G_{j_k}\}$ is a finite subcover of A in $\mathbb{G}$. Therefore, A is compact.

Example 6.1.6 The closed unit sphere $S[0, 1]$ centered at the origin $0 = \langle 0, 0, \ldots\rangle$ in l_2-space is not compact. The collection of open spheres $\left\{S\left(x, \frac{1}{2}\right) : x \in S[0, 1]\right\}$ is an open cover of $S[0, 1]$. Suppose that this open cover has a finite subcover

$$\left\{S\left(y_1, \frac{1}{2}\right), S\left(y_2, \frac{1}{2}\right), \ldots, S\left(y_k, \frac{1}{2}\right)\right\}$$

of $S[0, 1]$. Now consider the sequence $\langle x_n\rangle$, where $x_n = \langle 0, 0, \ldots, 1, 0, 0, \ldots\rangle$ is having 1 at the n^{th} place and 0 elsewhere. Then $\langle x_n\rangle$ is in $S[0, 1]$. Consequently, there is an i, $1 \leq i \leq k$,

such that $x_n \in S\left(y_i, \frac{1}{2}\right)$ for infinitely many values of n. Therefore, there exist distinct m and n in $\mathbb{Z}_+$ such that x_m and x_n are in $S\left(y_i, \frac{1}{2}\right)$. Then

$$d(x_m, x_n) \leq d(x_m, y_i) + d(y_i, x_n) < 1,$$

where d is the metric of the l_2-space. But $d(x_m, x_n) = \sqrt{2}$ for $m \neq n$.

Theorem 6.1.7 Any non-empty closed subset of a compact metric space is compact.

Proof. Let Y be a non-empty closed subset of a compact metric space (X, d). Let $\mathbb{G} = \{G_i\}_{i\in\Lambda}$, where G_i is open in Y, be an open cover of Y. Now for each G_i in $\mathbb{G}$, there exists an open set G_i^* in X such that $G_i^* \cap Y = G_i$. Now the collection of the open sets G_i^* obtained as above together with Y', the complement of Y in X, is an open cover of X. Note that Y' is open in X since Y is a closed set in X. Since X is compact, this open cover of X has a finite subcover of X. If Y' appears in this finite subcover, drop it. The remaining sets in the finite subcover contains Y. Then the collection of G_i corresponding to G_i^* in the finite subcover is clearly a finite subcover of Y in $\mathbb{G}$.

Example 6.1.8 Let $X = [0, 1]$ be with discrete metric. Then $X$is not compact. The subset $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$ is a closed set in X which is not a compact set.

Theorem 6.1.9 A compact subset of a metric space is closed and bounded.

Proof. Let (X, d) be a metric space, and A be a compact subset of X. Let p be any point in $X - A$. For each x in A, let $\varepsilon_x = \frac{d(p,x)}{2}$. Then the two open spheres $S(p, \varepsilon_x)$ and $S(x, \varepsilon_x)$ are disjoint, and the union of the collection $\mathbb{G}$ of all open spheres $S(x, \varepsilon_x)$, $x \in A$, obtained as above contains A. Since A is compact, by Theorem 6.1.5, there is a finite sub collection

$$\{S(x_1, \varepsilon_{x_1}),\ S(x_2, \varepsilon_{x_2}), \ldots, S(x_n, \varepsilon_{x_n})\}$$

of $\mathbb{G}$ which also covers A. Let $G = \bigcup_{i=1}^n S(x_i, \varepsilon_{x_i})$ and $G^* = \bigcap_{i=1}^n S(p, \varepsilon_{x_i})$. Then G and G^* are disjoint, and $A \subseteq G$. Therefore, G^* is an open set in X containing p, and which does not contain any point of A. Thus, p is not a limit point of A. Since p is arbitrary point of $X - A$, A is a closed set. To show that A is bounded, let x and y be two points in A. Then there exist i and j, $1 \leq i \leq n$, $1 \leq j \leq n$ such that $x \in S(x_i, \varepsilon_{x_i})$ and $y \in S(x_j, \varepsilon_{x_j})$. Then

$$\begin{aligned} d(x, y) &\leq d(x, x_i) + d(x_i, x_j) + d(x_j, y) \\ &\leq 2(\varepsilon_{x_1} + \varepsilon_{x_2} + \cdots + \varepsilon_{x_n}) + d(x_1, x_2) + d(x_1, x_3) + \cdots + d(x_{n-1}, x_n). \end{aligned}$$

This means that $\sup\{d(x, y) : x, y \in A\}$ is finite, and A is bounded.

The converse of the above theorem need not be true. Any infinite subset of a discrete metric space is closed and bounded but it is not compact.

Since a compact subset of a metric space is a closed set, the next two theorems follow from Theorem 6.1.9. It is however instructive to have another proof for them.

Theorem 6.1.10 Let (X, d) be a metric space, and p be a point of X. Let A be a compact subset of X not containing p. Then there exist two disjoint open sets in X one containing p and the other containing A.

Proof. For each x in A, there exist disjoint open sets U_x and V_x such that $x \in V_x$ and $p \in U_x$. Then A is covered by the collection $\{V_x\}$. Since A is compact, there exists a finite subcover $\{V_{x_1}, V_{x_2}, \ldots, V_{x_n}\}$ of A in $\{V_x\}$. Let $V = \bigcup_{i=1}^n V_{x_i}$, and $U = \bigcap_{i=1}^n U_{x_i}$. Then A is contained in V, $p \in U$, and V and U are disjoint open sets in X.

Theorem 6.1.11 In a metric space X, let A and B be two disjoint compact subsets. Then there exist two disjoint open sets in X one containing A and the other containing B.

Proof. By the preceding theorem, for each x in A, there exist two disjoint open sets U_x and V_x in X such that $x \in U_x$ and $B \subseteq V_x$. Then the collection $\{U_x\}$ is a cover of A, and so there is a finite subcover $\{U_{x_1}, U_{x_2}, \ldots, U_{x_n}\}$ of A. Let $U = \bigcup_{i=1}^{n} U_{x_i}$, and $V = \bigcap_{i=1}^{n} V_{x_i}$. Clearly U and V are disjoint open sets in X such that A is contained in U, and B is contained in V.

Theorem 6.1.12 Every compact metric space is totally bounded.

Proof. Let (X, d) be a compact metric space. Let $\varepsilon > 0$ be a real number. Then the collection of open spheres $\{S(x, \varepsilon) : x \in X\}$ is an open cover of X, and by the compactness of X, this open cover has a finite subcover of X, say, $\{S(x_1, \varepsilon), S(x_2, \varepsilon), \ldots, S(x_n, \varepsilon)\}$. Then the set $\{x_1, x_2, \ldots, x_n\}$ is a finite ε-net for X.

Corollary 6.1.13 Every compact metric space is separable.

Proof. Since a totally bounded metric space is separable by Theorem 2.17.6, the proof follows.

Theorem 6.1.14 Let A and B be compact subsets of a metric space X. Then $A \cup B$ is compact in X.

Proof. The proof is obvious.

EXERCISE

1. Show that the union of finite number of compact sets in a metric space is compact. Is the union of an infinite number of compact sets compact?

2. Let A be a subspace of a metric space X. Prove that the completion of A is compact if and only if A is totally bounded.

3. Show that the set $\left\{\frac{1}{n} : n \in \mathbb{Z}_+\right\} \cup \{0\}$ is a compact set in $\mathbb{R}$ directly from the definition. More generally, in any metric space (X, d), let $< x_n >$ be a sequence converging to $x \in X$. Then the set $\{x_n : n \in \mathbb{Z}_+\} \cup \{x\}$ is compact.

4. Construct a compact set in $\mathbb{R}$ whose set of limit points is denumerable.

5. Let A and B be disjoint sets in a metric space X such that A is compact, and B is closed. Show that there exists a $\delta > 0$ such that $d(p, q) > \delta$ if $p \in A$ and $q \in B$. Show that the conclusion may fail for two disjoint closed sets if neither is compact.

6. For each x in $]0, 1[$, let $I_x =]\frac{x}{2}, \frac{(x+1)}{2}[$. Show that the collection $\mathbb{G} = \{I_x : x \in]0, 1[\}$ is a cover of $]0, 1[$ which has no finite subcover of $]0, 1[$.

7. Add to cover $\mathbb{G}$ of the preceding example two appropriate members to form an open cover of $[0, 1]$. Show that this open cover has a finite subcover of $[0, 1]$.

8. Let X, Y and Z be metric spaces, and Y be compact. Let $f : X \to Y$ be a mapping, $g : Y \to Z$ be a one-to-one continuous mapping. Let $h(x) = g(f(x))$ for all x in X. Prove that f is uniformly continuous if h is uniformly continuous. Also prove that f

is continuous if h is continuous. Show that the compactness of Y cannot be omitted from the hypothesis even when X and Z are compact.

9. Let X be a compact metric space. Show that if $\{A_n\}$ is a countable collection of closed sets each of which has empty interior in X, then there exists a point in X which is not in any set A_n in the collection $\{A_n\}$.

10. Let X be a compact metric space, and $f : X \to X$ be a mapping such that $d(x, y) = d(f(x), f(y))$ for all x and y in X. Prove that f is surjective. Hint: If f is not surjective, then there is a point y in X such that $f(x) \neq y$ for all x in X. Consider the sequence $\langle y, f(y), f^2(y), \ldots\rangle$ in X.

11. Show that the Cantor set is compact.

12. Prove that the following statements are equivalent.

    (a) A metric space (X, d) is compact.

    (b) $X$is complete in every equivalent metric.

    (c) X is bounded in every equivalent metric.

13. Prove that if every closed sphere in a metric space X is compact, then X is complete and separable.

14. Let A be the set of all x in the closed interval [0,1] whose decimal expansion consists of only 4 and 7. Is A countable? Is A dense in [0,1] ? Is A compact? Is A perfect?

15. Is the set $A = \left\{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 - z^2 = 1\right\}$ compact in $\mathbb{R}^3$?

16. Prove that a subset A of l_1-space is compact if and only if A is bounded, and for each real number $\varepsilon > 0$, there exists an $n_0 \in \mathbb{Z}_+$ such that $\sum_{k=n_0}^{\infty} |x_k| < \varepsilon$ for all $x = \langle x_k\rangle$ in A.

17. Let X be a compact metric space. Let $\langle A_n\rangle$ be a sequence of non-empty closed set in X such that $A_{n+1} \subseteq A_n$ for all n. Show that the intersection $\bigcap_{n=1}^{\infty} A_n$ is non-empty.

18. Let (X, d) be a compact metric space, and $f : X \to Y$ be a mapping such that $d(f(x), f(y)) < d(x, y)$ for all x, y in X. Show that f has a fixed point. Is this fixed point unique?

19. Let X and Ybe metric spaces, and $\langle A_n\rangle$ be a nested sequence of compact subsets of X. Let $f : X \to Y$ be a continuous mapping. Show that $f\left(\bigcap_{n=1}^{\infty} A_n\right) = \bigcap_{n=1}^{\infty} f(A_n)$.

20. Let X be a compact metric space, and $f : X \to X$ be a continuous mapping. Show that there is a non-empty set A in X such that $f(A) = A$.

21. If closed and bounded sets in a metric space X are compact, then prove that X is complete.

22. Let X be a metric space. If for some real number $\varepsilon > 0$, every open sphere $S(x, \varepsilon)$, $x \in X$ has compact closure, then show that X is complete.

23. Give an example of a metric space X such that for each $x \in X$, there is a real number $\varepsilon > 0$ such that the open sphere $S(x, \varepsilon)$ has compact closure, and X is not complete.

24. Let (X, d) be a metric space, and$A$ be a subset of X. Let $\varepsilon > 0$ be a real number. Let $G(A, \varepsilon) = \bigcup_{x \in A} S(x, \varepsilon)$. Let $\mathbb{H}$ be the collection of all (non-empty) closed and bounded subsets of X. We define

$$D(A, B) = \inf \{\varepsilon : A \subseteq G(B, \varepsilon) \text{ and } B \subseteq G(A, \varepsilon)\}$$

for all $A, B \in \mathbb{H}$.

(i) Show that D is a metric on $\mathbb{H}$. D is called the Hausdorff metric.
(ii) If (X, d) is complete, then show that $(\mathbb{H}, D)$ is complete.
(iii) If (X, d) is totally bounded, then show that $(\mathbb{H}, D)$ is totally bounded.
(iv) If (X, d) is compact, then the space $(\mathbb{H}, D)$ of all non-empty closed and bounded subsets of X is compact.

25. Prove that a metric space is totally bounded if and only if its completion is compact.

26. Let (X, d) be a totally bounded metric space. Prove that for each real number $\varepsilon > 0$, there is a largest positive integers n such that it is possible to find n points $x_1, x_2, ..., x_n$ in X such that $d(x_i, x_j) \geq \varepsilon$ for $i \neq j$.

27. Let X be a metric space, and Y be a subset of X. If for each real number $\varepsilon > 0$, there is a finite set$\{x_1, x_2, ..., x_n\}$ in X such that $Y \subseteq \bigcup_{i=1}^{n} S(x_i, \varepsilon)$, then $Y$is totally bounded.

6.2 Sequential Compactness

The compactness is a very important property of metric spaces, and it has been investigated quite extensively. There are several conditions which are equivalent to compactness. At first sight these conditions look so different.

Definition 6.2.1 A metric space is said to be sequentially compact if every sequence in it has a convergent subsequence.

Examples 6.2.2

1. The real line $\mathbb{R}$ is not sequentially compact. The sequence $\langle n \rangle$ has no convergent subsequence.
2. An infinite discrete metric space (X, d) is not sequentially compact. From X we can extract a sequence of distinct points $\langle x_n \rangle$ which has no convergent subsequence.
3. In $\mathbb{R}^n$, the sequence $\langle x_n \rangle$, where $x_n = \langle n, 0, 0, \ldots \rangle$ has no convergent subsequence. Thus, $\mathbb{R}^n$ is not sequentially compact.
4. The interval $]0, 1[$ with usual metric is not sequentially compact. The sequence $\langle \frac{1}{n+1} \rangle$ in $]0, 1[$ has no convergent subsequence.
5. A finite metric space is sequentially compact since any sequence in such a metric space has a constant subsequence.
6. The closed interval $[a, b]$ in $\mathbb{R}$ is sequentially compact, since any sequence in $[a, b]$ is bounded, and thus has a convergent subsequence.
7. The rational numbers with usual metric do not form a sequentially compact metric space. The sequence $x_1 = 1$, $x_{n+1} = \frac{4+3x_n}{3+2x_n}$ for $n \geq 1$ has no convergent subsequence.
8. The interval $[a, \infty[$ in $\mathbb{R}$ is not sequentially compact.

The following theorem relates the notion of sequential compactness with the cluster point of a sequence.

Theorem 6.2.3 A metric space (X, d) is sequentially compact if and only if each sequence in X has a cluster point.

Proof. Assume that X is sequentially compact. Let $\langle x_n \rangle$ be a sequence in X. Then $\langle x_n \rangle$ has a convergent subsequence $\langle x_{n_k} \rangle$ which converges to a point x in X. Let $\varepsilon > 0$ be a real number. Then there exists $p \in \mathbb{Z}_+$ such that $d(x_{n_k}, x) < \varepsilon$ for all $k \geq p$. Now if m is any positive integer, then choose $q = \max\{n_p, n_m\}$. Then $q \geq p$, $n_q \geq m$, and $d(x_{n_q}, x) < \varepsilon$. Thus, x is a cluster point of $\langle x_n \rangle$. Conversely, let $\langle x_n \rangle$ be a sequence in X having a cluster point x in X. We can select a positive integer n_1 such that $d(x_{n_1}, x) < 1$. Now assume that there exists an n_k such that $d(x_{n_k}, x) < \frac{1}{k}$. Then, by the definition of a cluster point, there exists a positive integer $n_{k+1} > n_k$ such that $d(x_{n_{k+1}}, x) < \frac{1}{k+1}$. Then, by the principle of induction, there exists a subsequence $\langle x_{n_k} \rangle$ of $\langle x_n \rangle$ that converges to x. This proves that X is sequentially compact.

6.3 Bolzano-Weierstrass Property

In real line, a closed and bounded infinite set has a limit point. This is known as Bolzano-Weierstrass theorem. There are metric spaces in which this theorem is not true. For example, in the metric space $X =]0, 1]$ with usual metric, the closed and bounded set $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$ has no limit point in X. In an attempt to identify the class of metric spaces in which an infinite set has a limit point, it was found that this class coincides with the class of compact metric spaces. This motivates the following definition.

Definition 6.3.1 A metric space (X, d) is said to have Bolzano-Weierstrass property (BWP) if every infinite subset of X has a limit point in X.

The real line $\mathbb{R}$ does not have BWP. The subset of integers in $\mathbb{R}$ does not have any limit point in $\mathbb{R}$. The metric space $\mathcal{C}[a, b]$ in Example 2.1.10 does not have BWP since the set $\{f \in \mathcal{C}[a, b] : f(x) = n \text{ for all } x \in [a, b],\ n \in \mathbb{Z}\}$ does not have any limit point in $\mathcal{C}[a, b]$.

Theorem 6.3.2 A metric space is sequentially compact if and only if it has the Bolzano-Weierstrass property.

Proof. Let (X, d) be a sequentially compact metric space, and A be an infinite subset of X. We show that A has a limit point. Since A is infinite, we can extract a sequence $\langle x_n \rangle$ from A of distinct points. By our assumption, $\langle x_n \rangle$ has a convergent subsequence $\langle x_{n_k} \rangle$ which converges to $x \in X$. Let $\varepsilon > 0$ be a real number. Then there exists an $m \in \mathbb{Z}_+$ such that $d(x_{n_k}, x) < \varepsilon$ for all $k \geq m$. So that $x_{n_k} \in S(x, \varepsilon)$ for all $k \geq n$. Since all the points of $\langle x_{n_k} \rangle$ are distinct, the open sphere $S(x, \varepsilon)$ contains infinite number of points of A. Therefore, x is a limit point of A. Conversely, assume that every infinite subset of a metric space (X, d) has a limit point in X. Let $\langle x_n \rangle$ be a sequence in X. We show that $\langle x_n \rangle$ has a convergent subsequence. Let $A = \{x_n : n \in \mathbb{N}\}$ be the set of points of the sequence $\langle x_n \rangle$. If A is finite, then some element x_m of $\langle x_n \rangle$ will be repeated infinite number of times in the sequence, that is, $x_n = x_m$ for infinitely many values of n. Then there exists a constant subsequence $\langle x_m, x_m, \ldots \rangle$ of $\langle x_n \rangle$ which, of course, converges to x_m in X. If A is infinite, then, by our assumption, A has a limit point x in X. Then, by the definition of a limit point, for each positive integer k, there exists a positive integer n_k $(n_k > n_{k-1})$ such that $d(x_{n_k}, x) < \frac{1}{2^k}$.

Then $\langle x_{n_k} \rangle$ is a subsequence of $\langle x_n \rangle$ which converges to x. Therefore, X is sequentially compact.

Theorem 6.3.3 Every compact metric space has the Bolzano-Weierstrass property.

Proof. Let (X, d) be a compact metric space, and A be an infinite subset of X. Suppose that A has no limit point in X. Then no point of X is a limit point of A. Therefore, for each point x in X, there exists a real number $\varepsilon_x > 0$ such that the open sphere $S(x, \varepsilon_x)$ contains no point of A other than x. Then the collection $\mathbb{G}$ of such open spheres is an open cover of X. Since X is compact, there is finite subcollection

$$\{S(x_1, \varepsilon_{x_1}), S(x_2, \varepsilon_{x_2}), \ldots, S(x_n, \varepsilon_{x_n})\}$$

of $\mathbb{G}$ which covers X. Since each open sphere $S(x_i, \varepsilon_{x_i})$ contains at most one point of A, the union $\bigcup_{i=1}^{n} S(x_i, \varepsilon_{x_i})$ contains only finite number elements of A. But $A \subseteq X = \bigcup_{i=1}^{n} S(x_i, \varepsilon_{x_i})$ so that A is finite which contradicts our assumption that A is an infinite set. This completes the proof.

In view of the preceding two theorems, the three properties, viz, compactness, sequential compactness and BWP will be equivalent if we show that a sequentially compact metric space is compact. The following discussion is directed to achieve the later implication.

Consider the set $X = \left\{\frac{1}{n} : n \in \mathbb{Z}_+\right\}$ with usual metric. Then the set $\mathbb{G} = \left\{\left\{\frac{1}{n}\right\} : n \in \mathbb{Z}_+\right\}$ is an open cover of X. Let $\varepsilon > 0$ be a real number. Then there exists a m such that $\frac{1}{n} < \varepsilon$ for all $n \geq m$. Then the diameter of the set $A = \left\{\frac{1}{n} : n \geq m\right\}$ is less than ε. Evidently, the set A is not contained in any member of $\mathbb{G}$. Thus, for each real number $\varepsilon > 0$, there is a subset of X of diameter less than ε which is not contained in any member of $\mathbb{G}$. The metric space X under consideration is not compact. In compact metric spaces, we expect the converse of above situation to hold, and thus we define the following.

Definition 6.3.4 Let (X, d) be a metric space, and $\mathbb{G} = \{G_i\}$ be an open cover of X. A real number $\varepsilon > 0$ is called a Lebesgue number for the cover $\mathbb{G}$ if every subset of X with diameter less than ε is contained in some $G_i \in \mathbb{G}$.

Lemma 6.3.5 (Lebesgue) In a sequentially compact metric space, every open cover has a Lebesgue number.

Proof. Let (X, d) be a sequentially compact metric space, and $\mathbb{G} = \{G_i\}$ be an open cover of X which has no Lebesgue number. Then for each real number $\varepsilon > 0$, there exists a subset A_ε of X with diameter less than ε such that A is not contained in any member G_i of $\mathbb{G}$. Since for each real number $\varepsilon > 0$, the corresponding A_ε is not contained in any G_i in $\mathbb{G}$, A_ε must contain at least two elements. In particular, for each $n \in \mathbb{Z}_+$, the corresponding set $A_{\frac{1}{n}}$ contains at least two elements, and $\text{diam}(A_{\frac{1}{n}}) < \frac{1}{n}$. For each n, we choose an element x_n from $A_{\frac{1}{n}}$. Since X is sequentially compact, the sequence $\langle x_n \rangle$ has a convergent subsequence which converges to a point x in X. Now $x \in G_{i_0}$ for some G_{i_0} in $\mathbb{G}$. Since G_{i_0} is open, there exists a real number $\delta > 0$ such that $S(x, \delta)$ is contained in G_{i_0}. Let $S\left(x, \frac{\delta}{2}\right)$ be the concentric open sphere with radius $\frac{\delta}{2}$. Since a subsequence of $\langle x_n \rangle$ converges to x, $x_n \in S\left(x, \frac{\delta}{2}\right)$ for infinitely many values of n.

Let n_0 be one of these positive integers which is so large that $\frac{1}{n_0} < \frac{\delta}{2}$. Since $\text{diam}(A_{\frac{1}{n_0}}) < \frac{1}{n_0} < \frac{\delta}{2}$, $x_{n_0} \in S(x, \frac{\delta}{2})$, and $x_{n_0} \in A_{\frac{1}{n_0}}$. For any x' in $A_{\frac{1}{n_0}}$, we have

$$d(x, x') \leq d(x, x_{n_0}) + d(x_{n_0}, x') < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Therefore, $A_{\frac{1}{n_0}} \subseteq S(x,\delta) \subseteq G_{i_0}$. This contradicts the fact that $A_{\frac{1}{n_0}}$ is not contained in any of the members of $\mathbb{G}$, and completes the proof.

Theorem 6.3.6 Every sequentially compact metric space is totally bounded.

Proof. Let (X,d) be a sequentially compact metric space. Let $\varepsilon > 0$ be a real number. Choose a point x_1 in X. If $S(x_1,\varepsilon) = X$, then the set $\{x_1\}$ is a finite ε-net for X. If $S(x_1,\varepsilon) \neq X$, then select a point x_2 in X which is not in $S(x_1,\varepsilon)$. Now if $S(x_1,\varepsilon) \cup S(x_2,\varepsilon) = X$, then the set $\{x_1,x_2\}$ is a finite ε-net for X, otherwise we choose an element x_3 in the complement in X of the set $S(x_1,\varepsilon) \cup S(x_2,\varepsilon)$. If the above process ends after completing n steps, that is, $S(x_1,\varepsilon) \cup S(x_2,\varepsilon) \cup \ldots \cup S(x_n,\varepsilon) = X$, then the set $\{x_1,x_2,\ldots,x_n\}$ is a finite ε-net for X, and X will be totally bounded. Suppose that the above process continues indefinitely . Then the distance between any two elements of $\langle x_n \rangle$ will be greater than or equal to ε. Hence, the sequence $\langle x_n \rangle$ will not have any convergent subsequence. But this contradicts the fact that X is sequentially compact.

The following theorem gives the implication we were looking for.

Theorem 6.3.7 Every sequentially compact metric space is compact.

Proof. Let (X,d) be a sequentially compact metric space. Let $\mathbb{G} = \{G_i\}$ be an open cover of X. By Lemma 6.3.5, the cover $\mathbb{G}$ has a Lebesgue number ε. Let $\delta = \frac{\varepsilon}{3}$. By Theorem 6.3.6, there exists a finite δ-net $A = \{x_1,x_2,\ldots,x_n\}$ for X . Now the diameter $\operatorname{diam}(S(x_i,\delta)) \leq 2\delta = \frac{2\varepsilon}{3} < \varepsilon$. Therefore, for each $k = 1,2,\ldots,n$, by the definition of Lebesgue number, there exists a G_{i_k} in $\mathbb{G}$ such that $S(x_k,\delta) \subseteq G_{i_k}$. Since A is δ-net,

$$X = \bigcup_{k=1}^{n} S(x_k,\delta) \subseteq \bigcup_{k=1}^{n} G_{i_k} \subseteq X.$$

Thus, $\{G_{i_1}, G_{i_2}, \ldots, G_{i_n}\}$ is a finite subcover of X, and X is compact.

We have established the following.

Theorem 6.3.8 Let (X,d) be a metric space. The following statements are equivalent.

1. X is compact .
2. X is sequentially compact.
3. X has Bolzano-Weierstrass property.

Example 6.3.9 The metric space $\mathcal{C}[0,1]$ of continuous real valued functions defined on $[0,1]$ (Example 2.1.10) is not sequentially compact. Let $\langle f_n \rangle$ be a sequence in $\mathcal{C}[0,1]$ defined by

$$f_n(x) = \begin{cases} 0 & \text{if } 0 \leq x < \dfrac{1}{2^{n+1}}, \\ 2^{n+2}\left(x - \dfrac{1}{2^{n+1}}\right) & \text{if } \dfrac{1}{2^{n+1}} \leq x < \dfrac{3}{2^{n+2}}, \\ -2^{n+2}\left(x - \dfrac{1}{2^n}\right) & \text{if } \dfrac{3}{2^{n+2}} \leq x < \dfrac{1}{2^n}, \\ 0 & \text{if } \dfrac{1}{2^n} \leq x \leq 1. \end{cases}$$

It is obvious that $f_n \in \mathcal{C}[0,1]$, $d(f_n,f_m) = 1$ for $n \neq m$. Thus, the sequence $\langle f_n \rangle$ has no convergent subsequence.

EXERCISE

1. Show that the image of a sequentially compact metric space under a continuous mapping is sequentially compact.

2. Let (X, d) be a metric space, and Y be a non-empty subset of X. Then Y is said to be sequentially compact if Y is sequentially compact as a subspace of X. Show that every closed subset of a sequentially compact metric space is sequentially compact.

3. Show that the metric space $\mathcal{C}[0,1]$ with sup metric is not compact. Also show that a closed unit sphere $S[0,1]$ in $\mathcal{C}[0,1]$ is not compact.

4. Let $X =]0,1]$ and $G_n =]\frac{1}{n}, 1]$. Does a Lebesgue number exist for the open cover $\{G_n\}$ of X?

5. Prove that if a sequence in a compact metric space has a unique cluster point, then it is convergent.

6.4 Compactness and Finite Intersection Property

Theorem 6.4.1 A metric space is compact if and only if every family of closed sets with empty intersection has a finite subfamily with empty intersection.

Proof. Let X be a compact metric space, and $\mathbb{F} = \{F_i\}_{i\in\Lambda}$ be a family of closed sets such that $\bigcap_{i\in\Lambda} F_i = \emptyset$. Then $\bigcup_{i\in\Lambda} F_i' = X$, where $F_i' = X - F_i$. Then the collection $\{F_i'\}_{i\in\Lambda}$ is an open cover of X. By the compactness of X, this open cover has a finite subcover $\{F_{i_1}', F_{i_2}', \ldots, F_{i_n}'\}$ of X. Then $\bigcap_{j=1}^{n} F_{i_j}$ is empty, that is, $\{F_{i_1}, F_{i_2}, \ldots, F_{i_n}\}$ is a finite subfamily of $\mathbb{F}$ with empty intersection. Conversely, let $\{G_i\}_{i\in\Lambda}$ be an open cover of X. Then $\bigcap_{i\in\Lambda} G_i'$, where $G_i' = X - G_i$, is empty. By our assumption, there is a finite subfamily $\{G_{i_1}', G_{i_2}', \ldots, G_{i_k}'\}$ of the family $\{G_i' : i \in \Lambda\}$ with empty intersection, that is $\bigcup_{j=1}^{k} G_{i_j} = X$. Therefore, X is compact.

The above theorem can be restated in terms of the finite intersection property as follows.

Theorem 6.4.2 A metric space is compact if and only if every family of closed sets with the finite intersection property has non-empty intersection.

An immediate consequence of the above theorem is the following.

Theorem 6.4.3 (Cantor) A metric space X is compact if and only if every nested sequence $\langle F_n\rangle$ of non-empty closed subsets of X has the property that $\bigcap_{n=1}^{\infty} F_n$ is non-empty.

6.5 Compactness, Completeness and Total Boundedness

The following theorem shows that the class of compact metric spaces is a subclass of the class of complete metric spaces and also of the class of totally bounded metric spaces.

Theorem 6.5.1 A metric space is compact if and only if it is complete and totally bounded.

Proof. Let (X, d) be a metric space. First assume that X is compact. X is totally bounded by Theorem 6.1.12. Also, X is sequentially compact by Theorem 6.3.8. Then, every sequence in X has a convergent subsequence. In particular, every Cauchy sequence in X will have a convergent subsequence. Then every Cauchy sequence is convergent by Theorem 5.1.2. Therefore, X is complete. Conversely, let X be a complete and totally bounded metric space. To show X is compact, we show that X is sequentially compact. Consider an arbitrary sequence in X:

$$s_1 = \langle x_{11}, x_{12}, \ldots \rangle.$$

Since X is totally bounded, there exists a finite $\frac{1}{2}$-net $A_{\frac{1}{2}}$ for X. This means that there exists a point y_1 in $A_{\frac{1}{2}}$ such that $x_{1n} \in S(y_1, \frac{1}{2})$ for infinitely many values of n. These x_{1n}'s form a subsequence

$$s_2 = \langle x_{21}, x_{22}, \ldots \rangle$$

of s_1 such that $x_{2n} \in S(y_1, \frac{1}{2})$ for all n. Similarly, as above there exists a subsequence

$$s_3 = \langle x_{31}, x_{32}, \ldots \rangle$$

of s_2 such that $x_{3n} \in S(y_2, \frac{1}{3})$, where y_2 is a point in $\frac{1}{3}$-net $A_{\frac{1}{3}}$ for X, for all n. We continue in this way. Let s be the diagonal sequence.

$$s = \langle x_{11}, x_{22}, \ldots \rangle.$$

Since s_2 is a subsequence of s_1, s_3 is a subsequence of s_2, and so on, it follows that s is a subsequence of s_1. Now let $\varepsilon > 0$ be a real number. Then there exists a positive integer p such that $\frac{1}{p} < \frac{\varepsilon}{4}$. Now $x_{pn} \in S\left(y_{p-1}, \frac{1}{p}\right)$ for all n, and s_n is a subsequence of s_p for all $n \geq p$. Thus, it follows that $x_{nn} \in S\left(y_{p-1}, \frac{1}{p}\right)$ for all $n \geq p$. Therefore,

$$\begin{aligned} d(x_{nn}, x_{mm}) &\leq d(x_{nn}, x_{pp}) + d(x_{pp}, x_{mm}) \\ &< \frac{2}{p} + \frac{2}{p} = \frac{4}{p} < \varepsilon \text{ for all } n, m \geq p. \end{aligned}$$

Thus, s is a Cauchy subsequence of s_1 in X. Since X is complete, s is convergent. So s_1 has a convergent subsequence, and the proof is complete.

The above theorem shows that the property of total boundedness can be used to determine the compactness of metric spaces. There are certain metric spaces which occur as a closed subspaces of complete metric spaces, and to see whether these closed subspaces are compact subspaces or not it is sufficient to determine their total boundedness.

Corollary 6.5.2 A closed subset of a complete metric space is compact if and only if it is totally bounded.

Proof. Let X be complete, and A be a closed and totally bounded subspace of X. Then A is complete by Theorem 5.2.1, and compact by Theorem 6.5.1. The converse follows from Theorem 6.1.12.

Example 6.5.3 A closed unit sphere S in l_2-space is not compact. Since l_2-space is complete, S is complete. If S is compact, it would be totally bounded by Theorem 6.5.1. Since S is not totally bounded, it is not compact.

Corollary 6.5.4 Let X be a compact metric space. Then X is not the union of a countable number of closed sets none of which contains an open subset of X.

Proof. Since a compact metric space is complete, the proof follows from Corollary 5.5.9.

The Hilbert cube I^ω is the subset of l_2-space defined as

$$I^\omega = \left\{ \langle x_n \rangle \in l_2 : |x_n| < \frac{1}{n} \text{ for all } n \in \mathbb{Z}_+ \right\}.$$

Theorem 6.5.5 The Hilbert cube I^ω is a compact subset of the l_2-space.

Proof. First we show that I^ω is a closed set in l_2-space. Let $y = \langle y_n \rangle$ be a point in l_2-space which is not in I^ω. Then there exists an $n \in \mathbb{Z}_+$ such that $|y_n| > \frac{1}{n}$. Let $|y_n| = \alpha$ and $r = \alpha - \frac{1}{n}$. Let $x = \langle x_n \rangle$ be a point in I^ω. Then

$$d(x, y) = \left[\sum_{i=1}^{\infty} (x_i - y_i)^2 \right]^{\frac{1}{2}} \geq |x_n - y_n| > r.$$

Thus, I^ω does not intersect with $S(y, r)$, and consequently I^ω is a closed set in l_2-space. Therefore, I^ω is complete. By Theorem 6.5.1, it is enough to show that I^ω is totally bounded. Let $\varepsilon > 0$ be a real number. Since the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ is convergent, there exists a $m \in \mathbb{Z}_+$ such that $\sum_{k=m+1}^{\infty} \frac{1}{k^2} < \frac{\varepsilon^2}{2}$. Let $N_m = \{1, 2, \ldots, m\}$. For each $k \in \mathbb{Z}_+$, since the closed interval $\left[-\frac{1}{k}, \frac{1}{k}\right]$ is compact subset of $\mathbb{R}$, it is totally bounded. Therefore, for each $k \in \mathbb{Z}_+$ there exists a finite $\frac{\varepsilon}{\sqrt{2m}}$-net F_k for the interval $\left[-\frac{1}{k}, \frac{1}{k}\right]$. Let us define

$$F = \left\{ \langle x_n \rangle \in l_2 : x_k \in F_k \text{ for } k \in N_m \text{ and } x_k = 0 \text{ for } k \geq m + 1 \right\}.$$

It is obvious that F is a finite set. Let $y = \langle y_n \rangle$ be an element in I^ω. Let $k \in N_m$. Since F_k is a finite $\frac{\varepsilon}{\sqrt{2m}}$-net for $\left[-\frac{1}{k}, \frac{1}{k}\right]$, and $y_k \in \left[-\frac{1}{k}, \frac{1}{k}\right]$, there exists a point x_k in F_k such that $|x_k - y_k| < \frac{\varepsilon}{\sqrt{2m}}$. Let us take $x_k = 0$ for $k \geq m + 1$. Then the sequence $x = \langle x_1, x_2, \ldots, x_m, 0, 0, \ldots \rangle$ is in F, and

$$\begin{aligned} d(x, y) &= \left[\sum_{k=1}^{\infty} (x_k - y_k)^2 \right]^{\frac{1}{2}} = \left[\sum_{k=1}^{m} (x_k - y_k)^2 + \sum_{k=m+1}^{\infty} y_k^2 \right]^{\frac{1}{2}} \\ &\leq \left[m \frac{\varepsilon^2}{2m} + \sum_{k=m+1}^{\infty} \frac{1}{k^2} \right]^{\frac{1}{2}} < \varepsilon. \end{aligned}$$

This shows that F is a finite ε-net for I^ω, and the proof is complete.

Lemma 6.5.6 Let G and H be two disjoint non-empty open sets in a metric space X. Then $G \cap \overline{H} = \emptyset$ and $\overline{G} \cap H = \emptyset$.

Proof. The easy proof is left to the reader (see Theorem 7.1.7).

Theorem 6.5.7 Let X be a compact metric space such that every point of X is a limit point of X. Then X is uncountable.

Proof. The proof is divided into two steps.

Step-1. First we show that for any point x in X and any non-empty open set U in X, there exists a non-empty open set V contained in U such that $x \notin \overline{V}$. If x is in U, then

there exists a point y in U different from x since x is a limit point of X. If x is not in U, then there exists a point y in U since U is non-empty. Thus, in either case, there is a point y in U different from x. Then there exist disjoint open sets G and H such that $x \in G$ and $y \in H$. Let $V = H \cap U$. Then V is non-empty, and x is not in $\overline{V}$ by Lemma 6.5.6.

Step-2. Now we show that there is no surjection from $\mathbb{Z}_+$ to X. Let $f : \mathbb{Z}_+ \to X$ be a mapping defined by $f(n) = x_n$. By step-1, there exists a non-empty open set V_1 in X such that $x_1 \notin \overline{V}_1$. Again by step 1, there exists a non-empty open set $V_2 \subseteq V_1$ such that $x_2 \notin \overline{V}_2$. By induction, there exists a non-empty open set $V_n \subseteq V_{n-1}$ such that $x_n \notin \overline{V}_n$. Then the sequence $\langle \overline{V}_n \rangle$ is a nested sequence of non-empty closed sets. By Theorem 6.4.3, the intersection $\bigcap_{n=1}^{\infty} \overline{V}_n$ contains a point x_0. Now for any $n \in \mathbb{Z}_+$, the point x_0 can not be equal to x_n since $x_n \notin \overline{V}_n$.

The following theorem shows that the product of countable number of compact metric spaces is compact.

Theorem 6.5.8 Let $\{(X_i, d_i)\}_{i \in J}$ be a countable family of metric spaces, and (X, D) be their product metric space. Then (X, D) is compact if and only if each factor space (X_i, d_i), $i \in J$ is compact.

Proof. (X, D) is compact if and only if (X, D) is complete and totally bounded by Theorem 6.5.1 if and only if each (X_i, d_i) is complete and totally bounded by Theorem 2.17.11 and Theorem 5.2.3 if and only if each (X_i, d_i) is compact by Theorem 6.5.1.

6.6 Compact Sets in Real Line

As already remarked, Heine-Borel theorem identifies which sets in the real line are compact. These are precisely closed and bounded sets, and there are plenty of them.

Theorem 6.6.1 (Heine-Borel) A subset of the real line $\mathbb{R}$ is closed and bounded if and only if it is compact.

Proof. Any closed and bounded subset of the real line is a closed subset of some closed interval $[a, b]$ in the real line, and, by Theorem 6.1.7, it is sufficient to show that $[a, b]$ is compact. If $a = b$, then $[a, b]$ is compact. So, we assume that $a < b$. Let $\mathbb{G} = \{G_\alpha\}$ be a collection of open sets in $\mathbb{R}$ covering $[a, b]$. Let A be the set of points x in $[a, b]$ such that the closed interval $[a, x]$ can be covered by a finite number of open sets in $\mathbb{G}$. A is non-empty since a is in A. Since b is a an upper bound of A, A has its supremum. Let x_0 be the supremum of A. Then $a \leq x_0 \leq b$. Also note that A is an interval. Let $a < y < x$, and $x \in A$. Since $[a, x]$ can be covered by finite number of open sets in $\mathbb{G}$, $[a, y]$ can also be covered by finite number of open sets in $\mathbb{G}$. Thus, $y \in A$. Since $x_0 \in [a, b]$, there exists an open set $G_{\alpha'}$ in $\mathbb{G}$ containing x_0. Therefore, there exists a real number $\varepsilon > 0$ such that the open sphere $]x_0 - \varepsilon, x_0 + \varepsilon[$ is contained in $G_{\alpha'}$. If $x_0 = a$, then the interval $\left[a, a + \frac{\varepsilon}{2}\right] \subseteq]a - \varepsilon, a + \varepsilon[$. Thus, $\left[a, a + \frac{\varepsilon}{2}\right]$ is contained in $G_{\alpha'}$, and so $a + \frac{\varepsilon}{2}$ is in A, contradicting the fact that $x_0 = a$ is the supremum of A. Therefore, $a < x_0$. Next assume that $x_0 < b$. Since $x_0 - \frac{\varepsilon}{2}$ is not the supremum of A, $\left[a, x_0 - \frac{\varepsilon}{2}\right]$ can be covered by finite member of open sets, $G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}$ in $\mathbb{G}$. Then the closed interval $\left[a, x_0 + \frac{\varepsilon}{2}\right]$ can be covered by open sets $G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}$ together with $G_{\alpha'}$. This means that x_0 is not the supremum of A, a contradiction. So, $x_0 = b$. But then $\left[a, b - \frac{\varepsilon}{2}\right]$ can be covered by finite number of open sets in $\mathbb{G}$, and these together with $G_{\alpha'}$ will cover $[a, b]$. Therefore, $[a, b]$ is compact. The converse of the theorem follows from Theorem 6.1.9.

The Heine-Borel theorem above can be generalised to $\mathbb{R}^n$ as follows.

Theorem 6.6.2 A subset of the n-dimensional Euclidean plane $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. A compact subset of $\mathbb{R}^n$ is closed and bounded by Theorem 6.1.9. A closed and bounded subset of $\mathbb{R}^n$ is complete by Theorem 5.2.1, and totally bounded by Theorem 2.17.5, and therefore compact by Theorem 6.5.1.

Note that we could have adopted the same route as in the above theorem to establish Heine-Borel Theorem 6.6.1.

Example 6.6.3 The Cantor set C in $[0,1]$ is compact since it is a closed and bounded set in $\mathbb{R}$. Since every point of C is a limit point of C, by Theorem 6.5.7, the Cantor set C is uncountable. Since $\mathbb{R}$ contains C, $\mathbb{R}$ is uncountable. The above proof that $\mathbb{R}$ is uncountable is based only on the order properties of real number but not on any algebraic property of real numbers, that is, binary or ternary representation of real numbers, as used in Chapter 1.

Example 6.6.4 (Extended Real Line) Let us recall that the mapping $f : \mathbb{R} \to]-1,1[$ defined by $f(x) = x/(1+|x|)$ is a homeomorphism. Its inverse mapping is $g(x) = x/(1-|x|)$. Let J be the closed interval $[-1,1]$ with usual metric d. Let $\widetilde{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. We extend the mapping f to the bijection $\overline{f} = \widetilde{\mathbb{R}} \to J$ by putting $\overline{f}(-\infty) = -1$ and $\overline{f}(\infty) = 1$, and write again the inverse mapping $\overline{g}$ of $\overline{f}$. Now we can introduce a metric d' on $\widetilde{\mathbb{R}}$ defined as follows: Let u and v be any two points in $\widetilde{\mathbb{R}}$. Then there exist x and y in J such that $\overline{g}(x) = u$ and $\overline{g}(y) = v$ so that $\overline{f}(u) = x$ and $\overline{f}(v) = y$. Then the distance between u and v is defined as

$$d'(u,v) = d(x,y),$$

that is, $d'(u,v) = d(\overline{f}(u), \overline{f}(v)) = \left|\overline{f}(u) - \overline{f}(v)\right|$. Then d' is a metric on $\widetilde{\mathbb{R}}$ and the metric space $(\widetilde{\mathbb{R}}, d')$ is called the extended real line, usually denoted by $\widetilde{\mathbb{R}}$. When $\mathbb{R}$ is considered as a subspace of $\widetilde{\mathbb{R}}$, its subspace metric is different from the usual metric with respect to which $\mathbb{R}$ is called the real line. By the definition of d', it is easy to see that $d'(\infty, x) = \frac{1}{(1+|x|)}$ for $x \geq 0$, and $d'(-\infty, x) = \frac{1}{(1+|x|)}$ for $x < 0$.

We define an order relation on $\widetilde{\mathbb{R}}$ as follows: $x \leq y$ if $\overline{f}(x) \leq \overline{f}(y)$. On $\mathbb{R}$, this order relation coincides with the usual order on $\mathbb{R}$ introduced in Chapter 1. The order $\leq$ on $\widetilde{\mathbb{R}}$ satisfies all the properties that are satisfied by the usual order on $\mathbb{R}$ which can be stated purely in terms of the usual order (that is, excluding algebraic operations). Moreover, $-\infty < x < \infty$ for all x in $\mathbb{R}$. We call the elements of $\widetilde{\mathbb{R}}$ as extended real numbers, and the elements of $\mathbb{R}$ as real numbers. It may be noted that the order relation $\leq$ on $\widetilde{\mathbb{R}}$ is the one we encountered earlier in this text. In $\widetilde{\mathbb{R}}$, every set is bounded. The supremum, sup A, and infimum, inf A, of a subset A of $\widetilde{\mathbb{R}}$ always exist. The mapping $\overline{f} : (\widetilde{\mathbb{R}}, d') \to [-1,1]$ as defined above is clearly a homeomorphism so that the extended real line is compact since $[-1,1]$ is compact. The reader should note that this meaning of extended real line is distinct from the one in Section 4.9.

EXERCISE

1. Prove that a convergent sequence in $\mathbb{R}^n$ lies in a closed and bounded set in $\mathbb{R}^n$.

6.7 Compactness and Continuous Mappings

The following theorem shows that the property of compactness in metric spaces remains invariant under continuous mappings.

Theorem 6.7.1 Any continuous image of a compact metric space is compact.

Proof. Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a continuous mapping. We show that $f(X)$ is compact. Let $\mathbb{G} = \{G_i\}_{i\in\Lambda}$ be an open cover of $f(X)$ consisting of open sets G_i in $f(X)$. Now for each open set G_i in $\mathbb{G}$, there exists an open set G_i^* in Y such that $G_i^* \cap f(X) = G_i$. Since f is continuous, $f^{-1}(G_i^*)$ is an open set in X for each $i \in \Lambda$. The collection $\left\{f^{-1}(G_i^*)\right\}_{i\in\Lambda}$ is an open cover of X. Since X is compact, this open cover has a finite subcover of X, say,

$$\left\{f^{-1}(G_{i_1}^*),\, f^{-1}(G_{i_2}^*), \ldots, f^{-1}(G_{i_k}^*)\right\}.$$

Since $X = \bigcup_{j=1}^{k} f^{-1}(G_{i_j}^*)$, we have

$$f(X) = f(\bigcup_{j=1}^{k} f^{-1}(G_{i_j}^*)) = f(f^{-1}(\bigcup_{j=1}^{k} G_{i_j}^*)) \subseteq \bigcup_{j=1}^{k} G_{i_j}^*.$$

Therefore, $f(X) = \bigcup_{j=1}^{k} (G_{i_j}^* \cap f(X)) = \bigcup_{j=1}^{k} G_{i_j}$ and the proof is complete.

Corollary 6.7.2 Let (X, d_1) and (Y, d_2) be metric spaces, and A be a compact subset of X. Let $f : X \to Y$ be a continuous mapping. Then $f(A)$ is compact.

Corollary 6.7.3 The Euclidean space $\mathbb{R}^n$ is not compact.

Proof. Since the projection $p_1 : \mathbb{R}^n \to \mathbb{R}$ given by $p_1(x_1, x_2, \ldots, x_n) = x_1$ is continuous, and the real line $\mathbb{R}$ is not compact, it follows that $\mathbb{R}^n$ is not compact, for the compactness of $\mathbb{R}^n$ would imply the compactness of $\mathbb{R}$.

Corollary 6.7.4 If the product of two metric spaces is compact, then each factor space is compact.

Proof. Since the projection mappings are continuous, their ranges, which are factor spaces, are compact.

Since homeomorphisms and isometries are continuous mappings, the following corollary holds.

Corollary 6.7.5

1. Compactness is preserved under homeomorphisms.
2. Compactness is preserved under isometries.

The above corollary shows that compactness is a topological property.

Corollary 6.7.6 Let f be a continuous mapping of a compact metric space X into a metric space Y. Then the image $f(X)$ is a closed and bounded subset of Y.

Proof. The proof follows from Theorem 6.1.9 and Theorem 6.7.1.

Corollary 6.7.7 Let f be a continuous mapping of a compact metric space X into a metric space Y. Then f is a closed mapping.

Proof. If F is a closed subset of X, then F is compact. Therefore, $f(F)$ is compact, and thus a closed set in Y.

Theorem 6.7.8 Let f be a bijective and continuous mapping from a compact metric space (X, d_1) onto a metric space (Y, d_2). Then f is a homeomorphism.

Proof. It remains to show that f^{-1} is continuous. Let A be a closed set in X. Since X is compact, A is compact. Since f is continuous, $f(A)$ is compact. Therefore, $f(A)$ is a closed set in Y. Since f is bijective, $(f^{-1})^{-1}(A) = f(A)$, and the proof is complete.

In the above theorem, the compactness of X cannot be dropped.

Example 6.7.9 Consider $X = [0, 1[$ as a metric space with usual metric. X is not compact. Let C be the unit circle in the complex plane $\mathbb{C}$ with centre at the origin and with subspace metric of $\mathbb{C}$. Let $f : X \to C$ be defined by $f(x) = e^{2\pi ix}$ for all x in X. Then f is a bijective and continuous mapping such that f^{-1} is not continuous at $(1, 0) \in C$.

Theorem 6.7.10 Let f be a continuous real valued mapping on a metric space (X, d). Then f attains its supremum and infimum on every compact subset of X.

Proof. Let A be a compact subset of X. Then $f(A)$ is compact subset of $\mathbb{R}$. So $f(A)$ is a closed and bounded subset of $\mathbb{R}$. Also $f(A)$ is non empty. By order completeness of $\mathbb{R}$, $f(A)$ has both its supremum and infimum. Since $f(A)$ is closed, both supremum and infimum of $f(A)$ are in $f(A)$.

The above theorem is used to prove Rolle's Theorem which is at the heart of elementary calculus.

Theorem 6.7.11 (Rolle) Let f be a continuous real valued function defined on the closed interval $[a, b]$, $a < b$ in $\mathbb{R}$ such that $f(a) = f(b)$, and the derivative $f'(t)$ exists for all t in the open interval $]a, b[$. Then there exists a $t_0 \in]a, b[$ such that $f'(t_0) = 0$.

Proof. Since $[a, b]$ is compact, f attains it maximum value at a point $t_0 \in [a, b]$ by Theorem 6.7.10. If $f(t) > f(a)$ for some t in $]a, b[$, then $a < t_0 < b$. Therefore, by the differentiability of f at t_0

$$\lim_{h \to 0-} \frac{f(t_0 + h) - f(t_0)}{h} \geq 0$$

and $$\lim_{h \to 0+} \frac{f(t_0 + h) - f(t_0)}{h} \leq 0,$$

where the above limits represent the left hand and right hand derivatives at t_0, respectively. Consequently, $f'(t_0) = 0$. If $f(t) < f(a)$ for some t in $]a, b[$, then the minimum of f (which exists by the above theorem) is attained at same point s_0 in $]a, b[$. Then

$$\lim_{h \to 0-} \frac{f(s_0 + h) - f(s_0)}{h} \leq 0$$

and $$\lim_{h \to 0+} \frac{f(s_0 + h) - f(s_0)}{h} \geq 0.$$

Since $f'(s_0)$ exists, it follows that $f'(s_0) = 0$. Finally, if neither of the above two possibilities occur, then $f(t) = f(a) = f(b)$ for all t in $]a, b[$. In this case $f'(t) = 0$ for all t in $]a, b[$, and there surely exists a point u_0 in $]a, b[$ such that $f'(u_0) = 0$.

Corollary 6.7.12 Let X be a compact metric space, and Y be a metric space. Then the set $\mathcal{C}(X, Y)$ of continuous mappings from X into Y coincides with the set $\mathcal{C}_b(X, Y)$ of continuous and bounded mappings from X into Y.

Corollary 6.7.13 Let X be a compact metric space, and Y be a complete metric space. Then the metric space $\mathcal{C}(X, Y)$ of continuous mappings with sup metric from X into Y is complete.

Proof. The proof follows from Example 5.1.16 and Corollary 6.7.12.

A well known theorem in analysis of the real line states that a continuous real valued function defined on a closed interval $[a, b]$ is uniformly continuous. The property used in the proof is in fact the compactness of the domain set.

Corollary 6.7.14 Any continuous mapping from a compact metric space into a metric space is uniformly continuous.

Proof. Let (X, d_1) be a compact metric space, and (Y, d_2) be a metric space. Let $f : X \to Y$ be a continuous mapping. Let $\varepsilon > 0$ be a real number. Let x be a point in X, and consider the open sphere $S(f(x), \frac{\varepsilon}{2})$ in Y. Since f is continuous, $f^{-1}(S(f(x), \frac{\varepsilon}{2}))$ is an open set in X. Then the collection $\mathbb{G} = \left\{f^{-1}\left(S\left(f(x), \frac{\varepsilon}{2}\right)\right) : x \in X\right\}$ is an open cover of X. Since X is compact, it is sequentially compact Therefore, $\mathbb{G}$ has a Lebesgue number δ. Let x and x' be any points in X such that $d(x, x') < \delta$. Then the set $\{x, x'\}$ has diameter less than δ. Therefore, both x and x' are in $f^{-1}\left(S(f(x_0), \frac{\varepsilon}{2}\right) \in \mathbb{G}$. This means that $f(x)$ and $f(x')$ are both in $S(f(x_0), \frac{\varepsilon}{2})$, and so

$$d_2(f(x),\, f(x')) \leq d_2(f(x),\, f(x_0)) + d_2(f(x_0), f(x')) < \varepsilon.$$

This shows that f is uniformly continuous on X.

The following theorem indicates the condition under which the pointwise limit of a sequence of continuous real valued mappings is also the uniform limit (see Example 4.7.4).

Theorem 6.7.15 (Dini) Let $\langle f_n \rangle$ be a monotonically increasing sequence of continuous real valued mappings on a compact metric space (X, d). If $\langle f_n \rangle$ converges pointwise on X to a continuous real valued mapping f on X, then $\langle f_n \rangle$ converges uniformly to f on X.

Proof. Assume that $g_n = f - f_n$ for each n. Since $\langle f_n \rangle$ is monotonically increasing, $\langle g_n \rangle$ is monotonically decreasing, and $g_n(x) \geq 0$ for all x in X. Again, since $\langle f_n \rangle$ converges pointwise to f, $\langle g_n \rangle$ converges pointwise to zero mapping 0. We show that $\langle g_n \rangle$ converges to 0 uniformly on X. Let $\epsilon > 0$ be a real number. Since for any x in X, the sequence $\langle g_n(x) \rangle$ converges to 0, there exists an $m_x \in \mathbb{Z}_+$ such that $g_n(x) < \frac{\varepsilon}{2}$ for all $n \geq m_x$. Since g_{m_x} is continuous at x, there exists a real number $\delta_x > 0$ such that

$$y \in S(x, \delta_x) \text{ implies that } g_{m_x}(y) < \varepsilon.$$

Now the collection of open spheres $\{S(x, \delta_x) : x \in X\}$ obtained as above is an open cover of X. Since X is compact, this open cover has a finite subcover $\{S(x_1, \delta_{x_1}),\, S(x_2, \delta_{x_2}), \ldots,\, S(x_n, \delta_{x_n})\}$ of X. Assume that $m = \max\{m_{x_1}, m_{x_2}, \ldots, m_{x_n}\}$, and y be any point in X. Then there exists a k, $1 \leq k \leq n$ such that $y \in S(x_k, \delta_{x_k})$. Then $g_{m_{x_k}}(y) < \varepsilon$. Since $\langle g_n \rangle$ is monotonically decreasing sequence $g_m(y) \leq g_{m_{x_k}}(y) < \varepsilon$. So that $0 \leq g_n(y) < \varepsilon$ for all y in X and $n \geq m$. Therefore, $\langle g_n \rangle$ converges to 0 uniformly on X, and the proof is complete.

Theorem 6.7.16 Let A be a compact set in a metric space X. Then the sets $D_r(A)$ (see Example 2.4.2) form a fundamental system of neighbourhoods of A.

Proof. Let U be a neighbourhood of A. Then the mapping $f : X \to \mathbb{R}$ defined by $f(x) = d(x, X - U)$ is continuous on A by Theorem 4.8.3, and $f(x) > 0$ for all x in A. Since A is compact, there exists a point x_0 such that $f(x_0) = \inf_{x \in A} f(x)$ by Theorem 6.7.10.

But $f(x_0) = d(x_0, X - U) = r > 0$, that is, for any point y in $X - U$, $d(x, y) \geq r$ for all x in A. Thus, $D_r(A) \subseteq U$.

EXERCISE

1. Let (X, d) be a metric space which is not compact. Give an example of a real valued continuous and bounded mapping defined on X that (1) attains its supremum and infimum both (2) attains its infimum but not supremum (3) attains its supremum but not infimum (4) attains its neither supremum nor infimum.

2. Let A be a compact subset of a metric space (X, d), α be a real number, and $f : X \to \mathbb{R}$ be a mapping such that $|f(x) - f(y)| \leq \alpha\, d(x, y)$ for all x, y in A.

   Prove that for any sequence $\langle x_n \rangle$, the sequence $\langle f(x_n) \rangle$ has a convergent subsequence converging to $f(x_0)$ for some x_0 in A. Also prove that $f(A)$ is compact, and that f attains its infimum at some point in A.

3. Let (X, d) and (Y, ρ) be metric spaces, and A be a compact set in X. Let α be a real number, and $f : A \to Y$ be a mapping such that

   $$\rho(f(x),\ f(y)) \leq \alpha d(x, y).$$

   Show that $f(A)$ is compact.

4. Let A be a subset of $\mathcal{C}[0, 1]$ defined by $A = \{f \in \mathcal{C}[0, 1] : f : [0, 1] \to [0, 1]\}$. Show that A is a closed and bounded set, and it is not compact.

5. Let $]0, \infty[$ be with subspace metric of $\mathbb{R}$, and X be a compact metric space. Prove that for any continuous mapping $f : X \to]0, \infty[$, there exists a real number $\varepsilon > 0$ such that $f(x) \geq \varepsilon$ for all x in X.

6. Is the set $\left\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 - z^2 = 1\right\}$ compact in $\mathbb{R}^3$?

7. Let $f_n \in \mathcal{C}[0, 1]$ be defined by $f_n(t) = 1 - t^n$ for all $n \in \mathbb{Z}_+$. Show that the pointwise limit $f(t) = \lim_{n \to \infty} f_n(t)$ is not continuous on $[0, 1]$.

8. Let f be a continuous real valued function on the closed interval $[a, b]$ which is either strictly increasing or strictly decreasing on $[a, b]$. Then show that f is bijective, and f^{-1} is continuous on $f([a, b])$.

9. Let f be a strictly monotonic real valued function defined on $[a, b]$. Prove that f^{-1} is a continuous function of the subspace $f([a, b])$ of $\mathbb{R}$ onto $[a, b]$. Compare the questions 8 and 9.

10. Let $f : [a, b] \to \mathbb{R}$ be a continuous function such that for each $\xi \in]a, b[$, there exists $h \in]0, \xi[$ for which $f(\xi + h) > f(\xi)$. Show that $f(b) > f(a)$.

11. Prove the following statements which constitute the proof of the fundamental theorem of algebra: Every polynomial equation with complex coefficients and of degree $n \geq 1$ has at least one root in the field of complex numbers.

(a) Let $p : \mathbb{C} \to \mathbb{C}$ be a continuous function, and for each real number $M > 0$, there exists a real number $r > 0$ such that $|p(z)| \geq M$ whenever $|z| \geq r$, then there exists $\xi \in \mathbb{C}$ with $|p(\xi)| \leq |p(z)|$ for all $z \in \mathbb{C}$.

(b) Every polynomial $p(z) = a_0 + a_1 z + \cdots + a_n z^n$, where the coefficients $a_0, a_1, \ldots, a_n$ are in $\mathbb{C}$, which is not constant has the property (i).

(c) If p is a polynomial, and $\delta > 0$ is a real number such that $|p(0)| \leq |p(z)|$ whenever $|z| \leq \delta$, then $p(0) = 0$.

(d) If p is a polynomial which is not constant, then $p(\xi) = 0$ for some $\xi \in \mathbb{C}$.

12. Let $\mathbb{R}$ be the real line, and $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3 - 3x^2$. Show that the inverse image under f of a compact set is compact.

6.8 Countable Compactness

There is still another condition which is equivalent to the notion of compactness in metric spaces.

Definition 6.8.1 A metric space (X, d) is said to be countably compact if every countable open cover of X has a finite subcover of X.

It is straight forward from the above definition that a compact metric space is countably compact. The converse also holds, that is, a countably compact metric space is compact. First we establish the following.

Theorem 6.8.2 A metric space is countably compact if and only if every countable family of non-empty closed sets in X with the finite intersection property has non-empty intersection.

Proof. Assume that (X, d) is a countably compact metric space. Let $\mathbb{F}$ be a countable family of non-empty closed sets with the finite intersection property. If the intersection of this family is empty, then the family $\mathbb{F}'$ consisting of the complements of the members in $\mathbb{F}$ will be a countable open cover of X. By our assumption, this open cover $\mathbb{F}'$ will have a finite sub cover of X. Then the intersection of the elements in $\mathbb{F}$ whose complements are in the finite subcover is empty. This is a contradiction to the fact that $\mathbb{F}$ has the finite intersection property. Conversely, assume that $\mathbb{G}$ is a countable open cover of X. If for some G in $\mathbb{G}$, it is true that $G = X$, then there is nothing to prove. So let us assume that every member of $\mathbb{G}$ is a proper subset of X. Let us define $\mathbb{F}$ to be the family of complements in X of the sets in $\mathbb{G}$. Then $\mathbb{F}$ is a family of non-empty closed sets in X with empty intersection. Therefore, the union of the sets whose complements are in the finite subfamily is equal to X. The proof is now complete.

Theorem 6.8.3 A metric space is countably compact if and only if each sequence in X has a cluster point in X.

Proof. Let (X, d) be a countably compact metric space, and $\langle x_n \rangle$ be a sequence in X. For each $n \in \mathbb{Z}_+$, let $A_n = \{x_k : k \geqslant n\}$, and F_n be the closure of A_n. Then the family $\mathbb{F} = \{F_n : n \in \mathbb{Z}_+\}$ is a countable family of non-empty closed sets in X with the finite intersection property. By our assumption, the intersection $\cap\{F_n : n \in \mathbb{Z}_+\}$ is non-empty. Let x be a point in this intersection. Let N be a neighbourhood of x. For each $n \in \mathbb{Z}_+$, since $x \in F_n$, $A_n \cap N$ is non-empty. Consequently, there exists a positive integer $p \geq n$ such that $x_p \in A_n \cap N$. Therefore, x is cluster point of $\langle x_n \rangle$. Conversely, to prove that X

is countably compact, we prove the equivalent condition to the countable compactness in Theorem 6.8.2. So let $F = \{F_n : n \in \mathbb{Z}_+\}$ be a countable family of non-empty closed sets with finite intersection property. Since $S_n = \bigcap_{k=1}^{n} F_k$ is non-empty, there exists a point x_n in S_n. By our assumption, the sequence $\langle x_n \rangle$ so obtained has a cluster point x in X. Let m be any positive integer. Since x is a cluster point of $\langle x_n \rangle$, for each neighbourhood N of x, there exists positive integer $p \geq m$ such that $x_p \in N$. Then $S_p \cap N$ is non-empty. Therefore, $F_m \cap N$ is non-empty. Since F_m is a closed set, it follows that x is in F_m. Since m is arbitrary, x is in $\cap\{F_n : n \in \mathbb{Z}_+\}$, and the proof is complete.

We have proved the following.

Corollary 6.8.4 A metric space is countably compact if and only if it is sequentially compact.

Corollary 6.8.5 Every countably compact metric space is separable, complete and totally bounded.

6.9 Compact Sets in $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$

The metric spaces $\mathcal{C}_b(X, \mathbb{R})$ and $\mathcal{C}_b(X, \mathbb{C})$ of continuous and bounded real valued (Complex valued) mappings are very important from applications point of view. These metric spaces are shown to be complete in Example 5.1.16. Therefore, by Corollary 6.5.2, it follows that any closed subspace of $\mathcal{C}_b(X, \mathbb{R})$ or $\mathcal{C}_b(X, \mathbb{C})$ is compact if and only if it is totally bounded. From the practical point of view, determining the total boundedness of these subspaces can be a tedious task. This problem was overcome by obtaining an alternative criterion to determine the compactness of the closed subspaces. The proposed criterion is based upon the nature of individual elements in the closed subspace under consideration. If X is compact, then, by Corollary 6.7.12, the metric spaces $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$ of continuous mappings with sup metric coincide with $\mathcal{C}_b(X, \mathbb{R})$ and $\mathcal{C}_b(X, \mathbb{C})$, respectively, and we know that the elements in $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$ are uniformly continuous. We recall the definition of uniformly continuous mapping f in $\mathcal{C}(X, \mathbb{R})$ or $\mathcal{C}(X, \mathbb{C})$: f is uniformly continuous on X if for each real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that $d(x, y) < \delta$ implies that $|f(x) - f(y)| < \varepsilon$. The value of δ usually depends on the value of ε and the mapping f chosen. It may so happen that for a subset A of $\mathcal{C}(X, \mathbb{R})$ or $\mathcal{C}(X, \mathbb{C})$, a δ can be found such that for all x, y in X,

$$d(x, y) < \delta \text{ implies that } |f(x) - f(y)| < \varepsilon \text{ for all } f \text{ in } A,$$

then A is said to be equicontinuous. We make this explicit in the following definition.

Definition 6.9.1 Let (X, d) and (Y, ρ) be metric spaces, and $\mathcal{A}$ be a family of mappings from X into Y. Then $\mathcal{A}$ is said to equicontinuous if for each real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that for x, y in X

$$d(x, y) < \delta \text{ implies that } \rho(f(x),\, f(y)) < \varepsilon \text{ for all } f \in \mathcal{A}.$$

We observe that any member of an equicontinuous family is uniformly continuous.

Theorem 6.9.2 (Ascoli-Arzela) Let (X, d) be a compact metric space. Then a subspace of $\mathcal{C}(X, \mathbb{R})$ or $\mathcal{C}(X, \mathbb{C})$ is compact if and only if it bounded, equicontinuous and closed.

Proof. Let A be a subspace of $\mathcal{C}(X, \mathbb{R})$ or $\mathcal{C}(X, \mathbb{C})$. First assume that A is compact. Then A is closed and bounded. We show that A is equicontinuous. Let $\varepsilon > 0$ be a real number.

Since A is totally bounded, there exists a finite $\frac{\varepsilon}{3}$-net $A_1 = \{f_1, f_2, \ldots, f_n\}$ in A for A. Since X is compact, each f_k in A_1 is uniformly continuous. Thus, for each $k = 1, 2, \ldots, n$, there exists a real number $\delta_k > 0$ such that for all x, x' in X

$$d(x, x') < \delta_k \text{ implies that } |f_k(x) - f_k(x')| < \frac{\varepsilon}{3}.$$

Let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$. Then

$$d(x, x') < \delta \text{ implies that } |f_k(x) - f_k(x')| < \frac{\varepsilon}{3} \text{ for all } f_k \text{ in } A_1.$$

Now let f be any element in A. Since A_1 is a $\frac{\varepsilon}{3}$-net, there exists an element f_k in A_1 such that $d'(f, f_k) < \frac{\varepsilon}{3}$. Then for all x, x' in X

$d(x, x') < \delta$ implies that

$$\begin{aligned} |f(x) - f(x')| &\leq |f(x) - f_k(x)| + |f_k(x) - f_k(x')| + |f_k(x') - f(x')| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This means that A is equicontinuous. Conversely, we assume that A is bounded, closed and equicontinuous. We show the compactness of A by proving that it is sequentially compact. Since A is a closed subspace of the complete metric space $\mathcal{C}(X, \mathbb{R})$ (or $\mathcal{C}(X, \mathbb{C})$), it is complete. Thus, to show that A is sequentially compact, it is enough to show that every sequence in A has a Cauchy subsequence. Let

$$S_1 = \langle f_{11}, f_{12}, \ldots \rangle$$

be a sequence in A. Since A is bounded, there exists a real number k such that

$$|f(x)| \leq k \text{ for all } x \text{ in } X \text{ and } f \text{ in } A.$$

Since X is compact, it has a countable and dense set Y. Since Y is countable, its element can be arranged in the form of a sequence $\langle x_2, x_3, \ldots \rangle$. Now since a mapping f in A is bounded, the sequence $\langle f_{1j}(x_2) \rangle_{j=1}^{\infty}$ is a bounded sequence, and therefore it has a convergent subsequence. Let

$$S_2 = \langle f_{21}, f_{22}, \ldots \rangle$$

be a subsequence of S_1 such that the sequence $\langle f_{2j}(x_2) \rangle_{j=1}^{\infty}$ is convergent. Again the sequence $\langle f_{2j}(x_3) \rangle_{j=1}^{\infty}$ is bounded, and so will have a convergent subsequence. Let

$$S_3 = \langle f_{31}, f_{32}, \ldots \rangle$$

be a subsequence of S_2 such that the sequence $\langle f_{3j}(x_3) \rangle_{j=1}^{\infty}$ is convergent. By induction, we get the sequence S_{i+1} which is a subsequence of S_i and $\langle f_{i+1j}(x_{i+1}) \rangle_{j=1}^{\infty}$ is convergent. Let us consider the "diagonal" sequence

$$S = \langle f_{11}, f_{22}, f_{33}, \ldots \rangle.$$

Now S is a subsequence of S_1. Also for each point $x_j, j = 1, 2, \ldots$, the sequence $\langle f_{nn}(x_j) \rangle_{n=1}^{\infty}$ is a convergent sequence. We show that $\langle f_{nn} \rangle_{n=1}^{\infty}$ is a Cauchy sequence. Let $\varepsilon > 0$ be a real number. Since A is equicontinuous, there exists a real number $\delta > 0$ such that for x, x' in X

$$d(x, x') < \delta \text{ implies that } |f_{nn}(x) - f_{nn}(x')| < \frac{\varepsilon}{3} \text{ for all } n.$$

Since Y is dense, it follows that $\bigcup_{i=2}^{\infty} S(x_i, \delta) = X$. Thus, $\{S(x_i, \delta) : x_i \in Y\}$ is an open cover of X. Since X is compact, this open cover has a finite subcover

$$\{S(x_{i_1}, \delta), \ldots, S(x_{i_k}, \delta)\}$$

of X. Again since $\langle f_{nn}(x_j)\rangle_{n=1}^{\infty}$ is convergent for each $j = 2, 3, \ldots$, it follows that there exists an $n_0 \in \mathbb{Z}_+$ such that

$$|f_{nn}(x_j) - f_{mm}(x_j)| < \frac{\varepsilon}{3} \text{ for all } m, n \geq n_0, \text{ and } j = i_1, i_2, \ldots, i_k.$$

Now let x be any point in X. Then there exists an $i_j \in \{i_1, i_2, \ldots, i_k\}$ such that $x \in S(x_{i_j}, \delta)$, and so $d(x, x_{i_j}) < \delta$. Then

$$\begin{aligned} |f_{mm}(x) - f_{nn}(x)| &\leq \left|f_{mm}(x) - f_{mm}(x_{i_j})\right| + \left|f_{mm}(x_{i_j}) - f_{nn}(x_{i_j})\right| + \left|f_{nn}(x_{i_j}) - f_{nn}(x)\right| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \text{ for all } m, n \geq n_0. \end{aligned}$$

Thus, $\langle f_{nn}\rangle$ is a Cauchy subsequence of S_1, and the proof is complete.

As an application of Ascoli-Arzela theorem, we prove the Peano theorem that establishes the existence of a solution of the differential equation $\frac{dy}{dx} = f(x, y)$.

The following lemma reduces the problem of solving a differential equation to that of finding a solution of an integral equation.

Lemma 6.9.3 Let D be an open subset of the Euclidean plane $\mathbb{R}^2$, $(x_0, y_0) \in D$, and $f : D \to \mathbb{R}$ be a continuous mapping. Let $h > 0$ be a real number, and $I = [x_0 - h, x_0 + h]$. Let $\phi : I \to \mathbb{R}$ be a continuous function. Then $y = \phi(x)$ is a solution of the differential equation $\frac{dy}{dx} = f(x, y)$ passing through (x_0, y_0), that is, $y_0 = \phi(x_0)$, if and only if

$$\phi(x) = y_0 + \int_{x_0}^{x} f(t, \phi(t))dt \text{ for all } x \text{ in } I.$$

Proof. If $y = \phi(x)$ is a solution of $\frac{dy}{dx} = f(x, y)$ satisfying $\phi(x_0) = y_0$, then $\frac{d\phi}{dx} = f(x, \phi(x))$. For $x \in I$, integration between x_0 and x gives

$$\phi(x) = \phi(x_0) + \int_{x_0}^{x} f(t, \phi(t))dt = y_0 + \int_{x_0}^{x} f(t, \phi(t))dt \text{ for all } x \text{ in } I.$$

Conversely, if $\phi(x) = y_0 + \int_{x_0}^{x} f(t, \phi(t))dt$ for all x in I, then the continuity of f and the fundamental theorem of integral calculus yields $\phi'(x) = f(x, \phi(x))$, where $\phi'(x)$ denotes the derivative of $\phi(x)$ with resect to x, that is, $y = \phi(x)$ satisfies the equation $\frac{dy}{dx} = f(x, y)$ with $\phi(x_0) = y_0$.

In order to write the proof of the Peano theorem, we first describe the concept of a Cauchy polygon.

Let f be a continuous real valued mapping defined on a non-empty open subset D of the Euclidean plane $\mathbb{R}^2$, and $(x_0, y_0) \in D$. Since f is continuous, there exists an open sphere U centered at (x_0, y_0) and contained in D such that

$$|f(x, y)| \leq k \text{ for all } (x, y) \text{ in } U,$$

for some real number $k > 0$. Now we can choose a real number $\alpha > 0$ such that the rectangle

$$D_\alpha = \{(x, y) \in \mathbb{R}^2 : x \in [x_0 - \alpha,\ x_0 + \alpha] \text{ and } y \in [y_0 - \alpha k,\ y_0 + \alpha k]\}$$

is contained in U. Since D_α is compact, f is uniformly continuous on D_α. Then for each real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that for all $(x_1, y_1), (x_2, y_2) \in D_\alpha$

$$d((x_1, y_1), (x_2, y_2)) < \delta \text{ implies that } |f(x_1, y_1) - f(x_2, y_2)| < \frac{\varepsilon}{\alpha},$$

where d is the Euclidean distance on $\mathbb{R}^2$:

$$d((x_1, y_1), (x_2, y_2)) = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{\frac{1}{2}}.$$

We partition the closed interval $[x_0 - \alpha,\ x_0 + \alpha]$ as follows:

$$x_0 - \alpha = x_{-n} < x_{-n+1} < \cdots < x_{-1} < x_0 < x_1 < x_2 < \cdots < x_n = x_0 + \alpha,$$

and $\quad x_i - x_{i-1} < \delta(1 + k^2)^{\frac{1}{2}}$
for $i = -n + 1,\ -n + 2, \ldots, -1, 0, 1, 2, \ldots, n$. Now we draw a straight line in the plane $\mathbb{R}^2$ through (x_0, y_0) with slope $f(x_0, y_0)$, and suppose that this line intersects the line $x = x_1$ at the point (x_1, y_1). Since $|f(x_0, y_0)| \leq k$, the above line segment joining (x_0, y_0) and (x_1, y_1) lies inside the triangle bounded by the lines through (x_0, y_0) with slopes $\pm k$ and the line $x = x_0 + \alpha$ (Fig. 6.1). Next we draw a line through (x_1, y_1) with slope $f(x_1, y_1)$, and let this line meet the line $x = x_2$ at the point (x_2, y_2). We continue to draw such line segments till we reach the point (x_n, y_n) on the line $x = x_n = x_0 + \alpha$. We repeat the above process to the left of (x_0, y_0), until we reach the point (x_{-n}, y_{-n}) on the line $x = x_0 - \alpha$. Thus, we get a polygonal line joining the points (x_{-n}, y_{-n}) and (x_n, y_n). This polygonal line is called the Cauchy polygon ϕ of f. The graph of the Cauchy polygon ϕ of a function f for $n = 3$ is shown in Fig. 6.1.

The function ϕ is defined inductively as follows:

$\phi(x_0) = y_0$

$\phi(x) = \phi(x_{i-1}) + f(x_{i-1}, \phi(x_{i-1}))(x - x_{i-1})$ if $x_{i-1} < x \leq x_i$ and $i = 1, 2, ..., n$,

$\phi(x) = \phi(x_{i+1}) + f(x_{i+1}, \phi(x_{i+1}))(x - x_{i+1})$ if $x_i \leq x < x_{i+1}$ and $i = -1, -2, \ldots, -n$.

Let σ be the step function defined on $[x_0 - \alpha,\ x_0 + \alpha]$ by

$$\sigma(x) = \begin{cases} f(x_{i-1}, \phi(x_{i-1})) & \text{if } x_{i-1} < x \leq x_i \text{ and } i = 1, 2, \ldots, n, \\ f(x_{i+1}, \phi(x_{i+1})) & \text{if } x_i \leq x < x_{i+1} \text{ and } i = -1, -2, \ldots, -n, \\ f(x_0, y_0) & \text{if } x = x_0. \end{cases}$$

Then $|\sigma(x)| \leq k$ for all x in $[x_0 - \alpha, x_0 + \alpha]$. Also it follows that

$$\phi(x) = y_0 + \int_{x_0}^{x} \sigma(t)dt \text{ for all } x \in [x_0 - \alpha,\ x_0 + \alpha].$$

Now if x_1 and $x_2 \in [x_0 - \alpha,\ x_0 + \alpha]$, then

$$\begin{aligned} |\phi(x_1) - \phi(x_2)| &= \left| \int_{x_2}^{x_1} \sigma(t)\, dt \right| \\ &\leq |x_1 - x_2| \sup \{|\sigma(t)| : t \in [x_0 - \alpha,\ x_0 + \alpha]\} \\ &\leq k\, |x_1 - x_2|\,. \end{aligned}$$

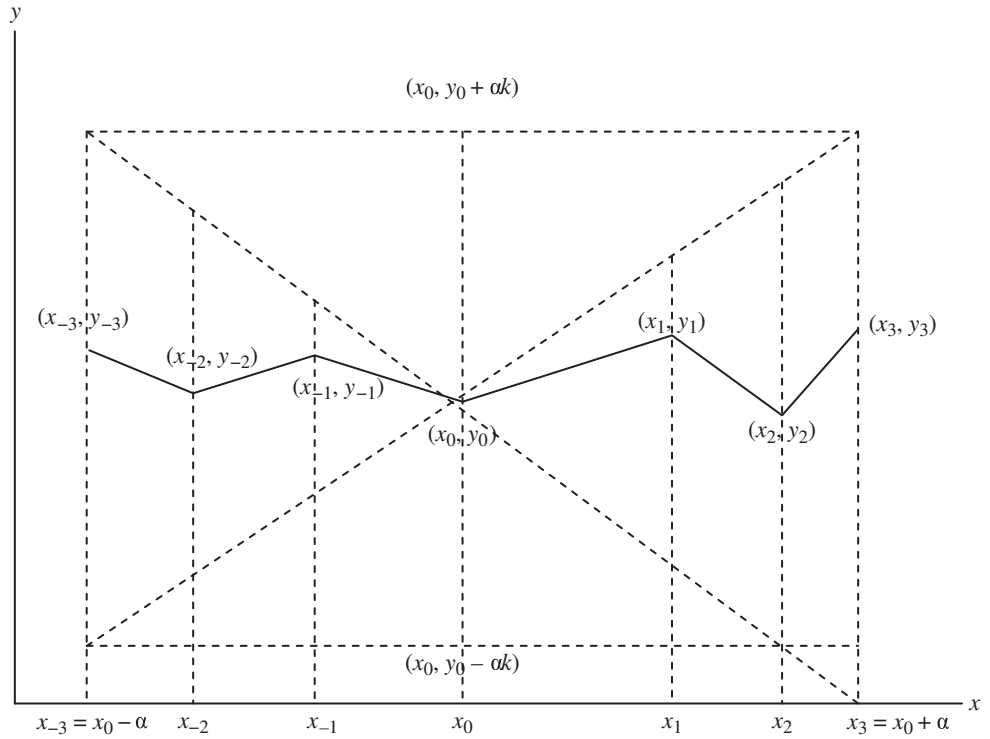


Fig. 6.1

Also for any $x \in [x_0 - \alpha,\, x_0 + \alpha]$,

$$|\phi(x) - y_0| = \left| \int_{x_0}^{x} \sigma(t)dt \right| \leq \alpha k.$$

Theorem 6.9.4 (Peano) Let f be a real valued continuous mapping defined on a non-empty open subset D of the Euclidean plane $\mathbb{R}^2$, and (x_0, y_0) be a point of D. Then there exist a real number $\alpha > 0$ and a function $\phi : [x_0 - \alpha,\, x_0 + \alpha] \to \mathbb{R}$ such that ϕ is a solution of the differential equation $\frac{dy}{dx} = f(x, y)$ which passes through the point (x_0, y_0).

Proof. By Lemma 6.9.3, it is sufficient to show that there exists a real number $\alpha > 0$ and a continuous function $\phi : [x_0 - \alpha,\, x_0 + \alpha] \to \mathbb{R}$ such that

$$\phi(x) = y_0 + \int_{x_0}^{x} f(t, \phi(t))dt \text{ for all } x \text{ in } [x_0 - \alpha,\, x_0 + \alpha]\,. \tag{6.1}$$

By the continuity of f, there exists an open sphere U with centre (x_0, y_0) contained in D such that for some real number k, $f(x, y) \leq k$ for all $(x, y) \in U$. We can choose a real number $\alpha > 0$ such that the rectangle

$$D_\alpha = \left\{(x, y) \in \mathbb{R}^2 : x \in [x_0 - \alpha,\, x_0 + \alpha] \text{ and } y \in [y_0 - \alpha k,\, y_0 + \alpha k]\right\}$$

is contained in U. Let $I = [x_0 - \alpha, x_0 + \alpha]$. Let A be the subset of $\mathcal{C}(I, \mathbb{R})$ containing all functions ϕ satisfying

$$|\phi(x) - y_0| \leq \alpha k \text{ for all } x \in I, \tag{6.2}$$

and

$$|\phi(x_1) - \phi(x_2)| \leq k\,|x_1 - x_2| \text{ for all } x_1, x_2 \in I. \tag{6.3}$$

Since the Cauchy polygon of f satisfies both the above conditions (6.2) and (6.3) on I, A is non-empty. Also the inequality (6.2) shows that A is bounded subset of $\mathcal{C}(I, \mathbb{R})$; and the inequality (6.3) shows that A is equicontinuous family in $\mathcal{C}((I, \mathbb{R})$. Therefore, by Ascoli-Arzela theorem, A is compact subset of $\mathcal{C}(I, \mathbb{R})$.

Now it is easily seen that for each ϕ in A, the mapping $t \to f(t, \phi(t))$ is a continuous function on I. We define a mapping $T : A \to A$ by

$$(T\phi)(x) = y_0 + \int_{x_0}^{x} f(t, \phi(t))dt \text{ for all } x \text{ in } I.$$

We show that $T\phi$ indeed belongs to A. Let $x_1, x_2 \in I$. Then

$$\begin{aligned} |(T\phi)(x_1) - (T\phi)(x_2)| &= \left| \int_{x_2}^{x_1} f(t,\, \phi(t))dt \right| \\ &\leq |x_2 - x_1| \sup \{f(t,\, \phi(t)) : t \in I\} \\ &\leq k\,|x_2 - x_1|\,. \end{aligned}$$

This shows that $T\phi$ is uniformly continuous and hence continuous. Also for $x \in I$

$$|(T\phi)(x) - y_0| = \left| \int_{x_0}^{x} f(t, \phi(t))dt \right| \leq \alpha k.$$

Thus, $T\phi \in A$. Since D_α is compact, f is uniformly continuous on D_α. Let $\varepsilon > 0$ be a real number. Then there exists a real number $\delta > 0$ such that for all (x_1, y_1) and (x_2, y_2) in D_α,

$$d((x_1, y_1), (x_2, y_2)) < \delta \text{ implies that } |f(x_1, y_1) - f(x_2, y_2)| < \frac{\varepsilon}{2\alpha}, \tag{6.4}$$

where d denotes the Euclidean distance on $\mathbb{R}^2$

$$d((x_1, y_1), (x_2, y_2)) = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{\frac{1}{2}}.$$

We show that T is continuous. Let $\phi_0 \in A$. Then for any $\phi \in A$,

$$\begin{aligned} |(T\phi)(x) - (T\phi_0)(x)| &= \left| \int_{x_0}^{x} \{f(t, \phi(t)) - f(t, \phi_0(t))\}\, dt \right| \\ &\leq |x - x_0| \sup \{|f(t, \phi(t)) - f(t, \phi_0(t))| \,:\, t \in I\}\,. \end{aligned}$$

Since $\|\phi - \phi_0\| < \delta$ implies that $|\phi(t) - \phi_0(t)| < \delta$ for all $t \in I$, from (6.4)

$$\|\phi - \phi_0\| < \delta \text{ implies that } |(T\phi)(x) - (T\phi_0)(x)| \leq |x - x_0|\, \frac{\varepsilon}{2\alpha}.$$

Thus, $\|T\phi - T\phi_0\| = \sup\{|(T\phi)(x) - (T\phi_0)(x)| : x \in I\} < \varepsilon$ which shows that T is a continuous mapping.

Let us define a mapping $F : A \to \mathbb{R}$ by $F(\phi) = \|T\phi - \phi\|$ for all ϕ in A. Then for all ϕ_1 and ϕ_2 in A,

$$\begin{aligned} |F(\phi_1) - F(\phi_2)| &= |\|T\phi_1 - \phi_1\| - \|T\phi_2 - \phi_2\|| \\ &\leq \|T\phi_1 - \phi_1 - (T\phi_2 - \phi_2)\| \\ &\leq \|T\phi_1 - T\phi_2\| + \|\phi_2 - \phi_1\| \end{aligned}$$

from which it follows that F is continuous on A. Since A is compact, F attains its infimum on A. Let $\phi_0 \in A$ be such that

$$F(\phi_0) = \inf\{F(\phi) : \phi \in A\}.$$

Since $F(\phi) = \|T\phi - \phi\| \geq 0$, $F(\phi_0) \geq 0$. If $F(\phi_0) = 0$, then it would follow that $T\phi_0 = \phi_0$, and hence from (1)

$$\phi_0(x) = (T\phi_0)x = y_0 + \int_{x_0}^{x} f(t, \phi_0(t))dt \text{ for all } x \text{ in } I.$$

Thus, ϕ_0 would satisfy (6.1). Therefore, to complete the proof, we show that $F(\phi_0) = 0$. Let $\varepsilon > 0$ be a real number. We construct a Cauchy polygon ϕ such that $F(\phi) < \varepsilon$. let δ be as in statement (6.4). Now consider the partition of the interval $[x_0 - \alpha, x_0 + \alpha]$

$$x_0 - \alpha = x_{-n} < x_{-n+1} < \cdots < x_{-1} < x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = x_0 + \alpha$$

such that

$$x_i - x_{i-1} < \frac{\delta}{(1+k^2)^{\frac{1}{2}}} \text{ for all } i = -n+1, -n+2, \ldots, n. \tag{6.5}$$

Now statements (6.3) and (6.5), for $i = 1, 2, \ldots, n$ and $x_{i-1} \leq t \leq x_i$, give

$$[(x_{i-1} - t)^2 + (\phi(x_{i-1}) - \phi(t))^2]^{\frac{1}{2}} \leq |x_{i-1} - t|\,(1+k^2)^{\frac{1}{2}} < \delta,$$

and so by statement (6.4)

$$|f(x_{i-1}, \phi(x_{i-1})) - f(t, \phi(t))| < \frac{\varepsilon}{2\alpha}.$$

Now we define the step function σ as

$$\sigma(x) = \begin{cases} f(x_{i-1}, \phi(x_{i-1})) & \text{if } x_{i-1} < x \leq x_i,\ i = 1, 2, \ldots, n, \\ f(x_{i+1}, \phi(x_{i+1})) & \text{if } x_i \leq x < x_{i+1}, i = -1, -2, \ldots, -n, \\ f(x_0, y_0) & \text{if } x = x_0. \end{cases}$$

Then $\phi(x) = y_0 + \int_{x_0}^{x} \sigma(t)dt$ for all x in I. Now let $x \in I$, and $x > x_0$ (the case when $x < x_0$

requires a very little modification). Then

$$\begin{aligned}
|\phi(x) - (T\phi)(x)| &= \left| \int_{x_0}^{x} \{\sigma(t) - f(t, \phi(t))\}\, dt \right| \\
&\leq \int_{x_0}^{x_0+\alpha} |\sigma(t) - f(t, \phi(t))|\, dt \\
&= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} |f(x_{i-1}, \phi(x_{i-1})) - f(t, \phi(t))|\, dt \\
&\leq \sum_{i=1}^{n} |x_i - x_{i-1}| \sup \{|f(x_{i-1}, \phi(x_{i-1}) - f(t, \phi(t)| \,:\, t \in [x_{i-1}, x_i]\} \\
&\leq \varepsilon \sum_{i=1}^{n} \frac{(x_i - x_{i-1})}{2\alpha} = \frac{\varepsilon}{2}.
\end{aligned}$$

Thus, $F(\phi) = \|T\phi - \phi\| = \sup \{|(T\phi)(x) - \phi(x)| \,:\, x \in I\} < \varepsilon$. Therefore, $F(\phi_0) = 0$, and the proof is complete.

EXERCISE

1. Let X be a compact metric space, and Y be a metric space. Let $F : X \times X \to Y$ be a continuous mapping. Let $f_y(x) = F(x, y)$. Show that $\mathcal{A} = \{f_y : y \in X\}$ is equicontinuous.

2. Let X be a convex and open subspace of $\mathbb{R}^n$. Let A be a family of differentiable functions from X into $\mathbb{R}^m$. Assume that there exists a real number $M > 0$ such that $\|f'(x)\| \leq M$, where f' denotes the derivative of f, for all f in $\mathcal{A}$ and for all $x \in X$. Show that $\mathcal{A}$ is equicontinuous.

3. Show that the collection of functions $\{\sin nx\}$ in $\mathcal{C}(\mathbb{R}, \mathbb{R})$ is not equicontinuous. Also show that the collection $\{f_n\}$, where $f_n(x) = nx$, $n \in \mathbb{Z}_+$, in $\mathcal{C}(\mathbb{R}, \mathbb{R})$ is not equicontinuous.

4. Show that the collection $\{f_n\}_{n \in \mathbb{Z}_+}$, where $f_n : [0, \infty[\to \mathbb{R}$ is defined by $f_n(x) = n \sin\left(\frac{x}{n}\right)$, is an equicontinuous family.

5. Let (X, d) be a metric space, and $f : X \to \mathbb{R}$ be a mapping. If there exist positive constants M and α with $0 < \alpha \leq 1$ such that

$$|f(x) - f(y)| \leq M\, [d(x, y)]^{\alpha} \text{ for all } x, y \text{ in } X,$$

   then f is a said to satisfy Holder's condition on X with constants M and α.

   Show that the family $\mathcal{C}$ of mappings satisfying Holder's condition with constants M and α is equicontinuous.

6. Give an example of an unbounded sequence $\langle f_n \rangle$ in $\mathcal{C}([0, 1], \mathbb{R})$ which is pointwise convergent to a function in $\mathcal{C}([0, 1], \mathbb{R})$.

7. Let $f_n(t) = \sin\left(\sqrt{t + 4n^2\pi^2}\right)$ for $t \geq 0$ and $n \in \mathbb{Z}_+$. Prove that the set $\{f_n : n \in \mathbb{Z}_+\}$ is a bounded and equicontinuous subset of $\mathcal{C}([0, \infty[, \mathbb{R})$. Prove also that the sequence $\langle f_n \rangle$ converges pointwise to 0 on $[0, \infty[$.

6.10 Locally Compact Metric Spaces

The n-dimensional Euclidean spaces $\mathbb{R}^n$ play an important role in the theory of integration. We know that these spaces are not compact, but closed and bounded sets in them are compact. Further, every point in $\mathbb{R}^n$ is contained in some closed and bounded and thus compact set. Speaking loosely, around every point there are "small" open sets which exhibit the property of compactness. We capture this idea in localisation of compactness in general metric spaces in the following definition.

Definition 6.10.1 A metric space X is said to be locally compact at a point x in X if there exists a neighbourhood of x whose closure is compact. X is called locally compact if it is locally compact at every point of X. A subset A of X is said to be locally compact if it is locally compact as a subspace of X.

Examples 6.10.2

1. The real line $\mathbb{R}$ is locally compact. For any point x in $\mathbb{R}$, there exists an open interval $]a, b[$ containing x whose closure $[a, b]$ is compact.

2. $\mathbb{R}^n$ is locally compact. For any point x in $\mathbb{R}^n$, there exists a neighbourhood of the form $]a_1, b_1[\times]a_2, b_2[\times \cdots \times]a_n, b_n[$ whose closure $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is compact.

3. The metric space $\mathbb{Q}$ of rational numbers with usual metric is not locally compact.

4. Any discrete metric space is locally compact. Note that any infinite discrete metric space is not compact.

Theorem 6.10.3 A compact metric space is locally compact.

Proof. Let X be a compact metric space. Then X is a neighbourhood of any point x in X, and the closure of this neighbourhood X is X itself which is compact.

The converse of this theorem is not true. An infinite discrete metric space is locally compact but it is not compact.

Theorem 6.10.4 A non-empty closed subset of a locally compact metric space is locally compact.

Proof. Let X be a locally compact metric space, and A be a non-empty closed subset of X. Let x be any point in A. Since X is locally compact, there exists an open set G containing x such that $\overline{G}$ is compact. Now, $G \cap A$ is open in A. Also $\overline{G} \cap A$, being the intersection of two closed sets in X, is a closed set in X. Since $\overline{G} \cap A$ is a subset of A, it is a closed set in A. Since $G \subseteq \overline{G}$, $cl(G \cap A) \subseteq cl(\overline{G} \cap A)$, where $cl(G \cap A)$ and $cl(\overline{G} \cap A)$ are closures of $G \cap A$ and $\overline{G} \cap A$ in A, respectively. Therefore, $cl(G \cap A) \subseteq \overline{G} \cap A$. Since $\overline{G}$ is compact subset of X, $\overline{G} \cap A$ is a compact subset of A. Thus, $cl(G \cap A)$ is compact, and the proof is complete.

The compactness remains invariant under continuous mappings but this is not true in the case of local compactness.

Example 6.10.5 Define $X = \{-1\} \cup]0, 1[$. Let X be given subspace metric of $\mathbb{R}$. Let Y be a subspace of $\mathbb{R}^2$ consisting of the point $(0, 0)$ and the points on the graph of the mapping $y = \sin \frac{1}{x}$, $0 < x < 1$. Let $f : X \to Y$ be the mapping defined by

$$f(-1) = (0, 0) \text{ and } f(x) = \left(x, \sin \frac{1}{x}\right) \text{ for all } x \text{ in }]0, 1[.$$

It is straightforward that f is continuous. It is easily seen that no open set containing $(0, 0)$ has compact closure. Thus, Y is not locally compact. X is clearly locally compact (Fig. 6.2).

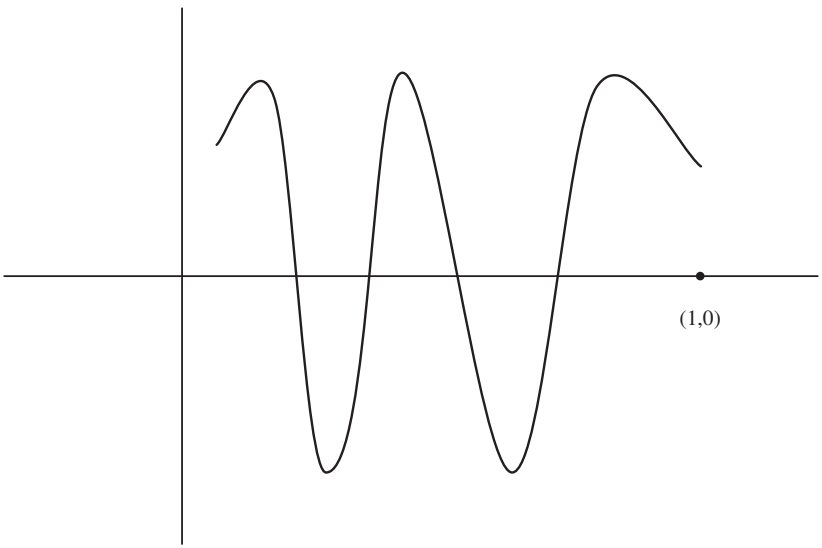


Fig. 6.2

The mapping f is not open, it maps the open set $\{-1\}$ in X to the set $\{(0, 0)\}$ which is not open in Y.

Theorem 6.10.6 The image of a locally compact metric space under open and continuous mapping is locally compact.

Proof. Let X be a locally compact metric space, and Y be a metric space. Let $f : X \to Y$ be an open and continuous mapping. We show that $f(X)$ is locally compact. For any point x in X, there exists an open set G containing x with compact closure. Then $f(G)$ is open set in Y. Since $f(G) \cap f(X) = f(G)$, $f(G)$ is open in $f(X)$, and it contains $f(x)$. Since f is continuous, $f(\overline{G})$ is compact. Therefore, $f(x) \in f(G) \subseteq f(\overline{G})$. Since $\overline{f(G)} \subseteq \overline{f(\overline{G})} = f(\overline{G})$, it follows that $\overline{f(G)}$ is compact, and the proof is complete.

Theorem 6.10.7 In a locally compact metric space X, every non-empty open set and every non-empty closed set is locally compact.

Proof. Let A be a non-empty open set in X, and x be a point in A. There exists an open set G containing x such that $\overline{G}$ is compact. Since G is open, there exists a real number $\varepsilon > 0$ such that the closed sphere $S[x, \varepsilon] \subseteq G$. Thus, $S[x, \varepsilon] \subseteq \overline{G}$, and so $S[x, \varepsilon]$ is compact by Theorem 6.1.7. Also, there exists a real number $\varepsilon > \varepsilon' > 0$ such that $S[x, \varepsilon']$ is contained in A, and it is compact by Theorem 6.1.7. Thus, A is locally compact.

Suppose that A is a non-empty closed set in X, and x be a point in A. Then there exists an open set G containing x such that $\overline{G}$ is compact. Then $\overline{G} \cap A$ is a closed set, and so $\overline{G \cap A} = \overline{\overline{G} \cap A} = \overline{G} \cap A$. Thus, $G \cap A$ is an open set containing x such that $\overline{G \cap A}$ is compact. This completes the proof.

Theorem 6.10.8 Let A be a compact set in a locally compact metric space. Then there exists a real number $r > 0$ such that the closure of the set $D_r(A) = \{x \in X : d(x, A) < r\}$ (see Example 2.4.2) is compact in X.

Proof. For each $x \in A$, there exists a compact neighbourhood V_x of x. Then the collection $\{V^0_{x_1} : x \in A\}$ is an open cover of A. Since A is compact, there exists a finite covering $\{V^0_{x_1}, V^0_{x_2}, \ldots, V^0_{x_n}\}$ of A. The set $V = \bigcup_{i=1}^{n} V_{x_i}$ is compact by Theorem 6.1.14, and is a neighbourhood of A. By Theorem 6.7.16, there exists $r > 0$ such that $D_r(A) \subseteq V$, and so $\overline{D_r(A)} \subseteq \overline{V} = V$. The proof is now complete.

Theorem 6.10.9 Let X be a locally compact metric space.Then the following statements are equivalent.

1. There exists a sequence of open sets $\langle U_n \rangle$ such that $\overline{U}_n$ is compact,

$$U_n \subseteq \overline{U}_n \subseteq U_{n+1} \quad \text{for every } n, \text{ and } X = \bigcup_{n=1}^{\infty} U_n.$$

2. X is a denumerable union of compact sets.

3. $X$is separable.

Proof. (1) implies (2). This is obvious since $X = \bigcup_{n=1}^{\infty} \overline{U}_n$, and $\overline{U}_n$ is compact.

(2) implies (3). Let X be the union of a sequence $\langle K_n \rangle$ of compact sets K_n. Then K_n is separable. Let A_n be a countable dense set in K_n, then $A = \bigcup_{n=1}^{\infty} A_n$ is countable and dense in X. Since $X = \bigcup_{n=1}^{\infty} K_n \subseteq \bigcup_{n=1}^{\infty} \overline{A}_n \subseteq \overline{A}$, X is separable.

(3) implies (1). Let X be a separable metric space. By Theorem 2.12.4, there exists a countable open base $\{V_n\}$ for X. Since X is locally compact, for each x in X, there is a compact neighbourhood W_x of x. Therefore, there exists an index n such that $x \in V_n \subseteq W_x$ by Theorem 2.12.2. Then the sets V_n such that $\overline{V}_n$ is compact constitute an open base for X. So we assume that each V_n in $\{V_n\}$ is with compact closure. Then we define U_n by induction as follows:

$$U_1 = V_1,$$
$$U_{n+1} = V_{n+1} \cup D_r(\overline{U}_n)$$

where $r > 0$ is a real number such that the closure of $D_r(\overline{U}_n)$ is compact. (This is possible by Theorem 6.10.8). Then the sequence $\langle U_n \rangle$ satisfies the statement (1).

Theorem 6.10.10 Let X_i, $i = 1, 2, \ldots, n$ be metric spaces, and $X = X_1 \times X_2 \times \cdots \times X_n$ be their product metric space. Then X is locally compact if and only if each factor space X_i is locally compact.

Proof. Let $x = (x_1, x_2, \ldots, x_n)$ be a point in X. Then in each X_i, there exists an open set G_i such that $x_i \in G_i$, and $\overline{G}_i$ is compact. Then $x \in G_1 \times \cdots \times G_n$, and $\overline{G} = \overline{G}_1 \times \overline{G}_2 \times \cdots \times \overline{G}_n$ by Theorem 2.15.5, and $\overline{G}$ is compact by Theorem 6.5.8 . Conversely, the projections

$\pi_i : X \to X_i$ are open and continuous onto mappings, and by Theorem 6.10.6, each X_i is locally compact.

Theorem 6.10.11 The l_2-space is not locally compact. If l_2-space were locally compact, then any point x in l_2-space would be contained in a compact neighbourhood. Then there would exists a real number $\varepsilon > 0$ such that the closed sphere $S[x, \varepsilon]$ was compact. Now, if $x = \langle x_i \rangle$, then the points $y_n = \langle x_1, x_2, \ldots, x_n + \varepsilon, x_{n+1}, \ldots \rangle$ are in $S[x, \varepsilon]$, and $d(y_n, y_m) = \sqrt{2}\varepsilon$ whenever $m \neq n$. Thus, the sequence $\langle y_n \rangle$ has no convergent subsequence which means that $S[x, \varepsilon]$ is not compact.

EXERCISE

1. Show that the subspace $\mathbb{Q}$ of rationals in the real line $\mathbb{R}$ is not locally compact.
2. Show that if every closed sphere in a metric space is compact, then the space is complete, locally compact and separable.
3. A metric space is locally compact if and only if there is an open base at each point all of whose sets have compact closures.

6.11 Peano Space Filling Curve

Peano constructed a continuous mapping from the closed interval $I = [0, 1]$ onto the square $I^2 = I \times I$ in $\mathbb{R}^2$. His construction of this function generates the same kind of excitement and also reverence for his genius as perhaps generated when one goes through the proof of the existence of a continuous everywhere and nowhere differentiable function (see Section 8.5). The construction changes one's naive geometrical intuition of continuous mappings so profoundly. However, the idea of construction of such a continuous mapping is rather simple.

Theorem 6.11.1 Let $I = [0, 1]$. Then there exists a continuous surjective mapping $f : I \to I^2$, where I is a subspace of $\mathbb{R}$, and I^2 is a subspace of $\mathbb{R}^2$.

Proof. We construct the mapping f as the limit of a sequence $\langle f_n \rangle$ of continuous mappings $f_n : I \to I^2$ as follows:

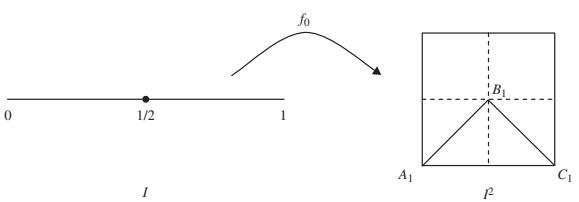


Fig. 6.3

Step-1. Define $f_0 : I \to I^2$ such that f_0 is a triangular path $A_1\ B_1\ C_1$ in I^2 (Fig. 6.3).

(The definition of f_0 can be written in a rigorous way but that is not required for our present purpose).

Step-2. The mapping $f_1 : I \to I^2$ is obtained as in Fig. 6.4 which is described as follows: we divide the sides of the square into four equal parts. Then f_1 consists of four triangular paths $A_1A_2A_3$, $A_3A_4A_5$, $A_5A_6A_7$, $A_7A_8A_9$ as shown, where each triangular path lies in a square of side $\frac{1}{2}$.

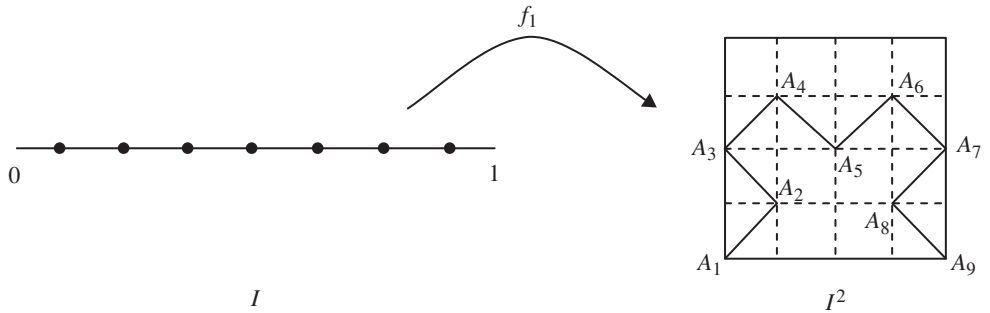


Fig. 6.4

Step-3. The mapping $f_2 : I \to I^2$ is obtained from f_1 by dividing each triangular path of f_1 into four triangular paths in the same way as the triangular paths of f_1 obtained from the triangular path of f_0. Note that each triangular path of f_2 is contained in a square of side $\frac{1}{2^2}$, and there are 4^2 triangular paths in all (Fig. 6.5).

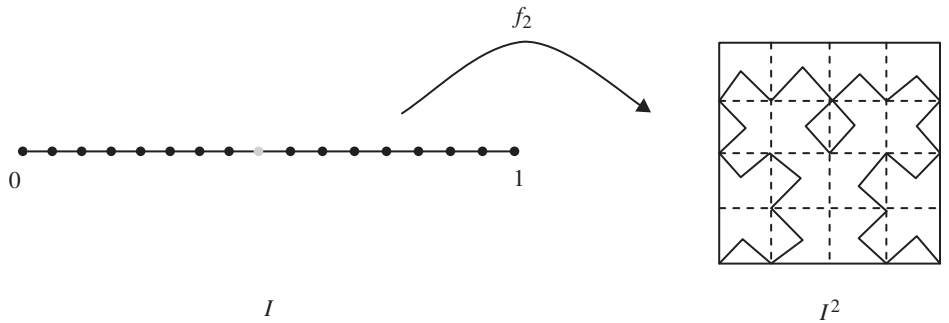


Fig. 6.5

Continuing in this way, the triangular paths of f_3 are shown in Fig. 6.6 which are 4^3 in number, and each such triangular path is contained in a square of side $\frac{1}{2^3}$.

In the n^{th} step, f_n contains 4^n triangular paths, and each triangular path is contained in a square of side $\frac{1}{2^n}$.

We show that $\langle f_n \rangle$ is a Cauchy sequence in the metric space $\mathcal{C}[I, I^2)$ of continuous mappings, where the metric on I^2 is the subspace metric of $\mathbb{R}^2$ given by

$$d((x_1, y_1), (x_2, y_2)) = \max\{|x_1 - x_2|, |y_1 - y_2|\},$$

and the metric on $\mathcal{C}(I, I^2)$ is the sup metric given by

$$\rho(f, g) = \sup\{d(f(t), g(t)) : t \in I\}.$$

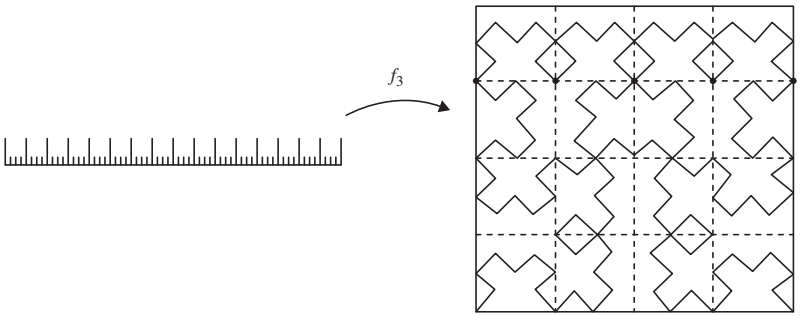


Fig. 6.6

Clearly I^2 is a closed set in $(\mathbb{R}^2, d)$. Since $(\mathbb{R}^2, d)$ is complete, I^2 is complete. Thus, $\mathcal{C}(I, I^2)$ is a complete metric space. Let us observe that when a triangular path in f_n contained in a square of edge $\frac{1}{2^n}$ is replaced by four triangular paths in f_{n+1}, these new triangular paths are contained in the same square. Hence,

$$d(f_n(t),\, f_{n+1}(t)) \leq \frac{1}{2^n} \text{ for all } t \in I.$$

Thus, $\rho(f_n, f_{n+1}) \leq \frac{1}{2^n}$. Now by the triangle inequality, for $n > m$,

$$\rho(f_n, f_m) \leq \frac{1}{2^m} + \frac{1}{2^{m+1}} + \cdots + \frac{1}{2^{n-1}} < \frac{2}{2^m}.$$

Therefore, $\langle f_n \rangle$ is a Cauchy sequence. Since $\mathcal{C}(I, I^2)$ is complete, $\langle f_n \rangle$ converges to a continuous mapping f in $\mathcal{C}(I, I^2)$. We prove that f is surjective. Let x be any point in I^2. For $n \in \mathbb{Z}_+$, x will be in a square of side $\frac{1}{2^n}$. Since f_n intersects with each such square, there exists a $t_0 \in I$ such that $d(f_n(t_0), x) < \frac{1}{2^n}$. Let $\varepsilon > 0$ be a real number. Then there exists a $p_1 \in \mathbb{Z}_+$ such that $\frac{1}{2^n} < \frac{\varepsilon}{2}$ for all $n \geq p_1$. Since $\langle f_n \rangle$ converges to f, there exists a $p_2 \in \mathbb{Z}_+$ such that $\rho(f_n, f) < \frac{\varepsilon}{2}$ for all $n \geq p_2$. Let $p = \max\{p_1, p_2\}$. Then

$$d(x, f(t_0)) \leq d(x, f_p(t_0)) + d(f_p(t_0), f(t_0)) < \varepsilon.$$

Therefore, $f(t_0) \in S(x, \varepsilon)$. Since ε is arbitrary, x is in the closure of $f(I)$. Since I is compact, $f(I)$ is compact, and is therefore a closed set. So x is in $f(I)$, and the proof is complete.

EXERCISE

1. For any n, show that there is a continuous surjective mapping $f : I \to I^n$.
2. Is there a continuous surjective mapping $f : I \to \mathbb{R}^2$?
3. Show that there is a continuous surjective mapping from $\mathbb{R}$ onto $\mathbb{R}^n$.

Chapter 7

Connectedness

If from the real line $\mathbb{R}$, the irrational numbers are dropped, it becomes a line with holes which is what we call the subspace of rational numbers. Intuitively, the absence of holes in the real line makes it a single piece while rationals are not a single piece. If we drop the origin $(0,0)$ from the 2-dimensional Euclidean space $\mathbb{R}^2$, it still remains a single piece while the removal of the y-axis will make two pieces of $\mathbb{R}^2$, *viz.* the right half plane and the left half plane. The property of being a single piece is very important in mathematical analysis, and is known as connectedness. All intervals in the real line $\mathbb{R}$ share this property. Thus, the real line is connected while the set of rational numbers is not. The metric spaces $\mathbb{R}^2$, $\mathbb{R}^2 - \{(0,0)\}$ are connected while $\mathbb{R}^2$ minus y-axis is not connected. It will be seen that the famous intermediate value theorem: If $f : [a,b] \to \mathbb{R}$ is a continuous function, and $f(a) \leq c \leq f(b)$, then there exists a x in $[a,b]$ such that $f(x) = c$, depends not only on the continuity of f but also on the property of connectedness of the interval $[a,b]$. Though the property of connectedness is quite intuitive, it requires some effort to formulate it precisely in mathematical terms, and so the reader is advised to have a little patience.

7.1 Separated Sets

Definition 7.1.1 Let (X,d) be a metric space. Two non-empty subsets A and B of X are said to separated if

$$A \cap \overline{B} = \emptyset, \text{ and } \overline{A} \cap B = \emptyset.$$

It follows from the above definition that separated sets are disjoint sets.

Examples 7.1.2

1. The intervals $]0,1[$ and $]1,2[$ are separated sets in the real line $\mathbb{R}$
2. The sets A and B defined by
$$A = \left\{(x,y) \in \mathbb{R}^2 : 0 < x < 1, \text{ and } y \text{ is arbitrary}\right\},$$
$$B = \left\{(x,y) \in \mathbb{R}^2 : 1 < x < 2, \text{ and } y \text{ is arbitrary}\right\}$$
are separated sets in $\mathbb{R}^2$.
3. Any two non-empty disjoint sets in a discrete metric space are separated sets.
4. In the complex plane $\mathbb{C}$, the sets
$$A = \{x + iy : x = 0 \text{ and } y \text{ is arbitrary}\},$$
$$B = \{x + iy : x \neq 0 \text{ and } y = 0\}$$
are not separated sets.

In a metric space, if A and B are separated sets, and P and Q are non-empty subsets of A and B, respectively, then P and Q are clearly separated sets.

The following results about separated sets are often useful.

Lemma 7.1.3 Let (X, d) be a metric space, and A, B and C be non-empty subsets of X such that A and B are separated sets, and also A and C are separated sets. Then A and $B \cup C$ are separated sets.

Proof. By our assumptions

$$A \cap \overline{B} = \phi \text{ and } \overline{A} \cap B = \emptyset,$$
$$A \cap \overline{C} = \phi \text{ and } \overline{A} \cap C = \emptyset.$$

Since $\overline{(B \cup C)} = \overline{B} \cup \overline{C}$,

$$A \cap \overline{(B \cup C)} = (A \cap \overline{B}) \cup (A \cap \overline{C}) = \emptyset$$

and

$$\overline{A} \cap (B \cup C) = (\overline{A} \cap B) \cup (\overline{A} \cap C) = \emptyset.$$

Theorem 7.1.4 Let (X, d) be a metric space, and A and B be separated sets in X. Then there exist two disjoint open sets U and V such that $A \subseteq U$, and $B \subseteq V$.

Proof. Let U be the set of all points x in X such that $d(x, A) < d(x, B)$, and V be the set of all x in X such that $d(x, A) > d(x, B)$. Since $d(x, B) > 0$ for all x in A and $d(x, A) = 0$ for all x in A, A is contained in U. Similarly, B is contained in V. Also, U and V are disjoint since $d(x, A) < d(x, B)$, and $d(x, A) > d(x, B)$ can not hold simultaneously. To show that U is an open set, let x be any point in U. Let $\delta = d(x, B) - d(x, A)$. Let y be any point in $S\left(x, \frac{\delta}{2}\right)$. Then from the triangle inequality, $d(y, A) < d(x, A) + \frac{\delta}{2}$. Also $d(x, B) < d(y, B) + \frac{\delta}{2}$, so that $d(y, B) + \frac{\delta}{2} > d(x, A) + \delta$, or $d(y, B) > d(y, A)$. Therefore, y is in U, and so $S\left(x, \frac{\delta}{2}\right)$ is contained in U. Thus, U is an open set. Similarly, V is an open set.

Corollary 7.1.5 Let (X, d) be a metric space, and A and B be disjoint closed sets in X. Then there exist disjoint open sets U and V in X such that $A \subseteq U$ and $B \subseteq V$.

Proof. The disjoint non-empty closed sets are separated sets.

Theorem 7.1.6 Let (X, d) be a metric space, and A and B be disjoint closed sets in X. Then there exist open sets U and V such that $A \subseteq U$, $B \subseteq V$, and $\overline{U} \cap \overline{V} = \emptyset$.

Proof. By Corollary 7.1.5, there exist disjoint open sets U and V such that $A \subseteq U$, and $B \subseteq V$. Then $X - U$ and A are disjoint closed sets. Thus, again, by Corollary 7.1.5, there exist disjoint open sets U^* and V^* such that $A \subseteq U^*$ and $X - U \subseteq V^*$. The sets U^* and V are the desired sets.

Theorem 7.1.7 Let (X, d) be a metric space. Then (a) if A and B are non-empty disjoint closed sets, then A and B are separated sets, (b) if A and B are non-empty disjoint open sets, then A and B are separated sets.

Proof. Part (a) is obvious. To prove part (b), assume that A and B are not separated sets. Then $A \cap \overline{B} \neq \emptyset$, or $\overline{A} \cap B \neq \emptyset$. Let $A \cap \overline{B} \neq \emptyset$, and x be a point in $A \cap \overline{B}$. Since A is open, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon)$ is contained in A. Again since $x \in \overline{B}$, and A and B are disjoint, $(S(x, \varepsilon) - \{x\}) \cap B \neq \phi$. Consequently, there exists a point $y \in (S(x, \varepsilon) - \{x\}) \cap B$. Then $y \in A \cap B$, a contradiction.

Theorem 7.1.8 Let (X, d) be a metric space, and A and B be non-empty subsets of X. If $d(A, B) > 0$, then A and B are separated sets.

Proof. Let $d(A, B) = \inf\{d(x, y) : x \in A \text{ and } y \in B\} = \alpha > 0$. Then $d(x, y) \geq \alpha$ for all x in A and y in B. This means that $A \cap B = \emptyset$, and no point of A is a limit point of B. Then $A \cap \overline{B} = \emptyset$. Similarly, $\overline{A} \cap B = \emptyset$.

EXERCISE

1. Let (X, d) be a metric space, and A and B be separated subsets of X. If $\emptyset \neq P \subseteq A$ and $\emptyset \neq Q \subseteq B$, then P and Q are separated sets.

2. Let (X, d) be a metric space, and $A, B \subseteq X$ such that $d(A, B) > 0$. Then A and B are separated sets. Show that the converse is not true.

3. Let (X, d) be a metric space, $G \subseteq X$, and $G = A \cup B$, where A and B are separated sets. Then (i) if G is open, then A and B are both open sets in X, (ii) if G is closed, then A and B are closed sets in X.

4. Let (X, d) be a metric space, and A and B be non-empty closed sets in X. Then $A - B$ and $B - A$ are separated sets.

7.2 Connected Metric Spaces

Definition 7.2.1 A metric space (X, d) is said to be disconnected if there exist separated sets A and B in X such that $X = A \cup B$. The pair (A, B) is called separation (or a disconnection) of X.

A metric space is said to be connected if it is not disconnected.

Examples 7.2.2

1. A metric space (X, d), where X is a singleton, is connected.
2. Any finite metric space having more than one points is disconnected.
3. A discrete metric space having more than one points is disconnected.
4. The real line is connected; any interval in the real line is connected
5. $\mathbb{R}^n$ and $\mathbb{C}^n$ are connected metric spaces.
6. The set of rationals $\mathbb{Q}$ with usual metric is disconnected. The pair (A, B), where $A = \{x \in \mathbb{Q} : x^2 < 2\}$ and $B = \{x \in \mathbb{Q} : x^2 > 2\}$, is a disconnection of $\mathbb{Q}$.
7. Let $X = \{(x, \sin \frac{1}{x}) : x \text{ is a positive real number}\}$, and d be the subspace metric of $\mathbb{R}^2$ on X. Then (X, d) is a connected metric space.

Theorem 7.2.3 Let (X, d) be a metric space. The following statements are equivalent.

1. (X, d) is disconnected.
2. X is the union of two non-empty disjoint closed sets.
3. X is the union of two non-empty disjoint open sets.
4. There exists a non-empty proper subset of X which is both open and closed.
5. There exists a continuous mapping $f : X \to \mathbb{R}$ such that $f(X)$ consists of exactly two points.

Proof. (1) implies (2). Let (A, B) be a separation of X. Then

$$\overline{A} \cap B = \emptyset,\ A \cap \overline{B} = \emptyset \text{ and } A \cup B = X.$$

Then $B \subseteq X - \overline{A} \subseteq X - A = B$. This means that $\bar{A} = A$. Similarly, $\overline{B} = B$. Thus, A and B are disjoint closed sets whose union is X.

(2) implies (3). Let A and B be non-empty disjoint closed sets such that $X = A \cup B$. Since $A = X - B$, and $B = X - A$, A and B are non-empty open sets, and X is their union.

(3) implies (4). Obvious.

(4) implies (5). Let A be a non-empty subset of X which is both open and closed. We define $f : X \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \in A, \\ 1 & \text{if } x \in X - A. \end{cases}$$

We show that f is a continuous mapping. Let G be an open set in $\mathbb{R}$. If 0 and 1 are both in G, then $f^{-1}(G) = X$ which is an open set. Similarly, if $0 \in G$ and $1 \notin G$, then $f^{-1}(G) = A$ which is an open set. Again if $1 \in G$ and $0 \notin G$, then $f^{-1}(G) = X - G$ which is an open set since G is a closed set. If none of 0 and 1 is in G, then $f^{-1}(G)$ is empty, and so, it is an open set. This shows that f is a continuous mapping.

(5) implies (1). Let $f : X \to \mathbb{R}$ be a continuous mapping such that $f(X) = \{a, b\}$, where $\{a, b\} \subseteq \mathbb{R}$. Let $f^{-1}(a) = A$, and $f^{-1}(b) = B$. Then A and B are non-empty disjoint sets, and X is their union. Since $\{a\}$ is a closed set in $\mathbb{R}$, A is a closed set. Similarly, B is a closed set. This means that (A, B) is a separation of X by Theorem 7.1.7.

In view of Statement (2) and (3) in the above theorem, the pair (A, B) where A and B are non-empty disjoint open (or closed) sets whose union is X, is also called a disconnection of X.

Corollary 7.2.4 A metric space is disconnected if and only if there exists a continuous mapping from X onto a metric space having two points.

7.3 Connected Sets

Definition 7.3.1 A subset of a metric space (X, d) is said to be disconnected if it is disconnected as a subspace of X. A subspace of a metric space is said to be connected if it is not disconnected. A subset of metric space is said to be connected if it is connected as a subspace of X.

By the above definition, it follows that a disconnected subset must at least have two elements.

Examples 7.3.2

1. In the real line $\mathbb{R}$, any finite set having at least two points is disconnected. The set of rationals, the set of irrationals and the set $\left\{\frac{1}{n} : n \in \mathbb{Z}_+\right\}$ are disconnected sets. The subset of $\mathbb{R}$ consisting of non-zero real numbers is disconnected .
2. The set $\mathbb{Q} \times \mathbb{Q}$ is disconnected subset of $\mathbb{R}^2$.
3. In a discrete metric space, any set having more than one points is disconnected.
4. Any interval in the real line is connected.

5. The set $\{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0\}$ in $\mathbb{R}^2$ is connected. The circle $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ in $\mathbb{R}^2$ is connected. Any line is connected set in $\mathbb{R}^2$. $\mathbb{R}^2 - \{(0, 0)\}$ is connected.
6. The set $S = \{(x, \sin \frac{1}{x}) \in \mathbb{R}^2 : x > 0\}$ is a connected subset of R^2.

Theorem 7.3.3 Let (X, d) be a metric space, and A be a non-empty subset of X. Then the following statements are equivalent.

1. A is disconnected.
2. There exists two non-empty closed sets G and H in X such that $G \cap A$ are $H \cap A$ are non-empty, $G \cap H \cap A$ is empty, and $A \subseteq G \cup H$.
3. There exist two non-empty open sets G and H in X such that $G \cap A$ are $H \cap A$ are non-empty, $G \cap H \cap A$ is empty, and $A \subseteq G \cup H$.

Proof. (1) implies (2). Since A is disconnected, by Theorem 7.2.3, there exists non-empty disjoint closed sets G^* and H^* in A such that A is the union of G^* and H^*. Then there exist closed sets G and H in X such that $G^* = G \cap A$ and $H^* = H \cap A$. Then G and H are required sets.

(2) implies (3). Let G and Hbe non-empty closed sets in X satisfying our assumption in (2). Then $G \cap A$ and $H \cap A$ are non-empty disjoint closed sets in A whose union is A. Then $G \cap A$ and $H \cap A$ are also open sets in A. Therefore, there exist open sets U and V in X such that $G \cap A = U \cap A$ and $H \cap A = V \cap A$. Hence, $A \subseteq U \cup V$, U and V are non-empty, intersect with A, and $U \cap V \cap A$ is empty.

(3) implies (1). If G and H are open sets satisfying the conditions in (3), then $G \cap A$ and $H \cap A$ are non-empty disjoint open sets in A whose union is A, and the proof follows from Theorem 7.2.3.

Theorem 7.3.4 Let (X, d) be a metric space. A subset Y of X is disconnected if and only if Y is expressible as the union of two separated sets.

Proof. If Y is disconnected, then there is a disconnection (A, B) of Y. Then $cl(A) \cap B = \emptyset$, and $A \cap cl(B) = \emptyset$, where $cl(A)$ and $cl(B)$ denote the closures of A and B in Y, respectively. Then $cl(A) = \overline{A} \cap Y$, where $\overline{A}$ is the closure of A in X. Similarly, $cl(B) = \overline{B} \cap Y$. Then $cl(A) \cap B = (\overline{A} \cap Y) \cap B = \overline{A} \cap B$. Similarly, $A \cap cl(B) = A \cap \overline{B}$. Therefore, $A \cap \overline{B} = \emptyset$ and $\overline{A} \cap B = \emptyset$. Conversely, if A and B are separated sets in X such that $Y = A \cup B$, then $cl(A) \cap B = (\overline{A} \cap Y) \cap B = \overline{A} \cap B = \emptyset$, where $cl(A)$ and $\overline{A}$ are as defined earlier in the proof. Similarly, $A \cap cl(B) = \emptyset$, and the proof is complete.

From the above theorem, it follows that the subsets A and B of Y are separated in Y if and only if A and B are separated sets in X.

The following result is frequently used, and follows from the above theorem.

Corollary 7.3.5 Let X be a metric space, and Y be a subspace of X. A subset U of Y is disconnected in Y if and only if U is disconnected in X.

Theorem 7.3.6 Let $\mathcal{A} = \{C_i\}$ be a collection of connected subsets of a metric space X, and for every A and B in $\mathcal{A}$, there exists a finite chain $C_0, C_1, C_2, \ldots, C_n$ in $\mathcal{A}$ such that $C_0 = A, \quad C_n = B$ and $C_j \cap C_{j+1} \neq \emptyset$ for $1 \leq j \leq n-1$. Then $\bigcup \{C_i : C_i \in \mathcal{A}\}$ is connected.

Proof. Let $A_0 = \bigcup \{C_j : C_j \in \mathcal{A}\}$. Suppose on the contrary that A_0 is disconnected, and $A_0 = B \cup C$, where B and C are non-empty disjoint open sets in A_0. Since B and C are

non-empty, there exist G and H in $\mathcal{A}$ such that G intersects B, and H intersects C. Then G must be entirely contained in B, for if $G \cap B$ and $G \cap C$ are non-empty, then $G \cap B$ and $G \cap C$ are disjoint open sets in G such that $G = (G \cap B) \cup (G \cap C)$, and this means that G is disconnected, a contradiction. Similarly, H is entirely containing in C. By our assumption, there exists a chain $C_0, C_1, \ldots, C_n$ in $\mathcal{A}$ such that $G = C_0$, $H = C_n$ and $C_j \cap C_{j+1} \neq \emptyset$ for all j. As above, each C_j is entirely contained either in B or in C. Now $C_0 \subseteq B$. Assuming $C_j \subseteq B$, where $j < n$. Since $C_j \cap C_{j+1} \neq \emptyset$, it follows that $C_{j+1} \subseteq B$. So, by induction, C_n is contained in B which is a contradiction, and the proof is complete.

The following theorem is a consequence of the above theorem.

Theorem 7.3.7 Let $\mathcal{A} = \{C_i\}$ be a collection of connected subsets of a metric space X, and $C_{i_0} \in \mathcal{A}$ such that $C_{i_0} \cap C_j \neq \emptyset$ for all C_j in $\mathcal{A}$. Then $\cup\{C_j : C_j \in \mathcal{A}\}$ is connected.

Proof. Let $A_0 = \bigcup\{C_j : C_j \in \mathcal{A}\}$. Suppose on the contrary that $A_0 = B \cup C$, where B and C are non-empty open sets in A_0 such that $B \cap C = \emptyset$. Then each $C_j$in $\mathcal{A}$ must lie entirely in B or C, for otherwise C_j is disconnected. If C_{i_0} lies in B, then $C_j \subseteq B$ for all C_j in $\mathcal{A}$. Then $A_0 \subseteq B$, and so C is empty, a contradiction.

Corollary 7.3.8 Let $\mathcal{A}$ be a collection of connected subsets of a metric space X such that $A \cap B \neq \emptyset$ for all A and B in $\mathcal{A}$. Then $\cup\{A : A \in \mathcal{A}\}$ is connected.

Proof. For any A in $\mathcal{A}$, it follows that $A \cap B \neq \emptyset$ for all sets B in $\mathcal{A}$. Now apply the preceding theorem.

Corollary 7.3.9 Let $\mathcal{A}$ be a collection of connected subsets in a metric space X such that $\bigcap\{C_j : C_j \in \mathcal{A}\}$ is non-empty. Then $\cup\{C_j : C_j \in \mathcal{A}\}$ is connected.

Proof. Let $A_0 = \cap\{C_j : C_j \in \mathcal{A}\}$. Let A be any member of $\mathcal{A}$. Then A intersects with A_0, and hence with every C_j in $\mathcal{A}$.

Corollary 7.3.10 Let $\langle C_i \rangle_{i=1}^n$ be a sequence of connected sets in a metric space X such that $C_i \cap C_{i+1} \neq \emptyset$ for all $i = 1, 2, \ldots, n-1$. Then $\bigcup_{i-1}^n C_i$ is connected.

Proof. The proof follows from Theorem 7.3.6.

Theorem 7.3.11 If A is a connected set in a metric space X, then any set B in X such that $A \subseteq B \subseteq \bar{A}$ is connected. In particular, $\overline{A}$ is connected.

Proof. Suppose that C and D are two non-empty disjoint open sets in B such that $B = C \cup D$. As A is dense in B, $A \cap C$ and $A \cap D$ are both non-empty open sets in A. Also $A = (A \cap C) \cup (A \cap D)$, and $(A \cap C) \cap (A \cap D) = \phi$. Thus, A is not connected, a contradiction.

Example 7.3.12 Let (X, d) be a connected metric space in which d is not bounded. Then, for any point x in X and a real number $r > 0$, the set $A = \{y \in X : d(x, y) = r\}$ is non-empty. Suppose, if possible, the set A is empty. Then the set $B = \{y \in X : d(x, y) < r\}$ and $C = \{y \in X : d(x, y) > r\}$ are open sets in X, B is a non-empty since x is in B, and C is non-empty since d is not bounded. Thus, (B, C) is a disconnection of X, a contradiction.

Theorem 7.3.13 Let (X, d) be a metric space. If a connected subset A of X is contained in the union of two separated sets in X, then A is contained in one of them.

Proof. Let P and Q be separated sets in X such that $A \subseteq P \cup Q$. If $P \cap A$ or $Q \cap A$ is empty, then the assertion is obvious. So we assume that $P \cap A \neq \emptyset$ and $Q \cap A \neq \emptyset$. As $P \cap A \subseteq P$ and $Q \cap A \subseteq Q$, it follows that $P \cap A$ and $Q \cap A$ are separated sets, and A is their union Therefore, A is disconnected, a contradiction.

Theorem 7.3.14 Let (X, d) be a metric space, and A be a non-empty subset of X. Then A is connected if and only if every pair of points in A lies in a connected subset of A.

Proof. Let A be a disconnected set. Then there exist non-empty closed sets P and Q such that $A = P \cup Q$. Since P and Q are non-empty, let $x \in P$ and $y \in Q$. Then x and y are in A. Then, by our assumption, there exists a connected subset B such that $x, y \in B$ and $B \subseteq A$. Then B is entirely contained either in P, or in Q. If we assume that $B \subseteq P$, then y is in P, a contradiction. The converse is obvious.

Theorem 7.3.15 Let A be a subset of a metric space X. If B is a connected subset of X such that $A \cap B$ and $(X - A) \cap B$ are not empty, then $bd(A) \cap B$ is not empty. In particular, if X is connected, then any non-empty proper subset A of X has at least one boundary point.

Proof. Suppose that $bd(A) \cap B$ is empty. Denote $X - A$ by A'. Since $X = A^0 \cup (A')^0 \cup bd(A)$, B is the union of $A^0 \cap B$ and $(A')^0 \cap B$ both of which are open sets in B, and non-empty by assumption, for a point in $A \cap B$ must belong to $A^0 \cap B$ since $bd(A) \cap B = \emptyset$, and similarly for $A' \cap B$. As $(A^0 \cap B) \cap (A')^0 \cap B = \emptyset$, this means that B is not connected, a contradiction.

EXERCISE

1. Give an example of a connected set which is not compact.
2. Let A and B be two non-empty subsets in a metric space. If A and B are closed sets in X, and $A \cup B$ and $A \cap B$ are connected, then A and B are connected. Give an example in the real line that the assumption that both A and B are closed can not be deleted.
3. Let X be a connected metric space, and Y be a metric space. Let $f : X \to Y$ be a continuous mapping. Show that the graph of f
$$G_f = \{(x, f(x)) : x \in X\}$$
is connected in the product metric space $X \times Y$.
4. In the plane $\mathbb{R}^2$ with distance $d((x_1, x_2), (y_1, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$, let Y be the set $\{(x_1, x_2) : x_1 = 0 \text{ and } 0 \leq x_2 \leq 1, \text{or } 0 \leq x_1 \leq 1 \text{ and } x_2 = 0\}$. Show that every open sphere in Y is connected but the closure of an open sphere $S((x_1, x_2), r)$ may not be the closed sphere $S[(x_1, x_2), r)]$.
5. Prove that the metric space $\mathcal{C}[a, b]$ with sup metric is connected.
6. If A and B are connected subsets of a metric space X, show that $A \times B$ is connected in the product metric space $X \times X$.
7. If in a metric space X, the closure of every open sphere $S(x, r)$ is the closed sphere $S[x, r]$, then show that every open sphere in X is connected.
8. In a connected metric space, show that the boundary of a non-empty set is non-empty.
9. Prove that a connected metric space having more than one point is uncountable.
10. Let X be the set of points in $\mathbb{R}^2$ which have at least one rational co-ordinate. Show that with induced metric from $\mathbb{R}^2$, X is a connected metric space.

7.4 Connectedness and Continuous Mappings

The connectedness is preserved under continuous mappings. This fact is proved in the next theorem.

Theorem 7.4.1 The continuous image of a connected set is connected.

Proof. Let X and Y be two metric spaces, and S be a connected subset of X. Let $f : X \to Y$ be a continuous mapping. We show that $f(S)$ is connected subset of Y. Suppose, on the contrary that $f(S)$ is disconnected. Then there exists non-empty disjoint open sets U and V in $f(S)$ such that $f(S) = U \cup V$. Then there are open sets U^* and V^* in Y such that

$$U = U^* \cap f(S) \text{ and } V = V^* \cap f(S).$$

Since f is continuous, $f^{-1}(U^*)$ and $f^{-1}(V^*)$ are both open sets in X. Then $f^{-1}(U^*) \cap S$ and $f^{-1}(V^*) \cap S$ are both open in S. Since $f(S) = U \cup V$, we have

$$S \subseteq f^{-1}(f(S)) = f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V) \subseteq f^{-1}(U^*) \cup f^{-1}(V^*).$$

This means that

$$S = (f^{-1}(U^*) \cup f^{-1}(V^*)) \cap S = (f^{-1}(U^*) \cap S) \cup (f^{-1}(V^*) \cap S).$$

Further, $f^{-1}(U^*) \cap S$ and $f^{-1}(V^*) \cap S$ are non-empty sets. To see this, observe that U and V are non-empty, and $f(S) = U \cap V$. Then $f^{-1}(U) \cap S$ and $f^{-1}(V) \cap S$ are both non-empty, and now use the inclusions $f^{-1}(U) \subseteq f^{-1}(U^*)$, and $f^{-1}(V) \subseteq f^{-1}(V^*)$ to establish the claim. Finally, we show that the sets $f^{-1}(U^*) \cap S$ and $f^{-1}(V^*) \cap S$ are disjoint. Let $x \in (f^{-1}(U^*) \cap S) \cap (f^{-1}(V^*) \cap S)$. Then $f(x) \in U^*$, $f(x) \in V^*$, and $f(x) \in f(S)$. This means that $f(x)$ is in $U \cap V$, a contradiction. This concludes that S is disconnected, and the proof is complete.

The special case of the above is the following corollary.

Corollary 7.4.2 The continuous image of a connected metric space is connected.

Corollary 7.4.3 A metric space X is connected if and only if every continuous mapping f from X into any two point metric space $Y = \{a, b\}$ is a constant mapping.

Proof. If f is not constant, then it is a surjection. Since $\{a\}$ and $\{b\}$ are open sets in Y, $(f^{-1}(\{a\}), f^{-1}(\{b\}))$ is a disconnection of X, a contradiction.

Conversely, if X is disconnected, then there exists a disconnection (A, B) of X in terms of open sets. The mapping f defined by $f(A) = \{a\}$ and $f(B) = \{b\}$ is continuous, a contradiction.

Corollary 7.4.4 The connectedness is preserved under homeomorphisms.

Corollary 7.4.5 The connectedness is preserved under isometries.

Example 7.4.6 The unit sphere $S^n = \left\{x = (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : \|x\| = 1\right\}$ is connected subset of $\mathbb{R}^n$. To show this, let $S^n_+ = \{(x_1, x_2, \ldots, x_{n+1}) \in S^n : x_{n+1} \geq 0\}$, be the upper hemisphere, and $S^n_- = \{(x_1, x_2, \ldots, x_{n+1}) \in S^n : x_{n+1} \leq 0\}$ be the lower hemisphere. Let D_n be the closed unit sphere $\{u \in \mathbb{R}^n : \|u\| \leq 1\}$ in $\mathbb{R}^n$. Note that D_n is a convex, and hence a connected set in $\mathbb{R}^n$. We show that the mapping $f_+ : D_n \to S^n_+$ defined by

$$f_+(u) = (u_1, u_2, \ldots, u_n, \sqrt{1 - \|u\|^2})$$

for all $u = (u_1, u_2, \ldots, u_n)$ in D_n is a homeomorphism. f_+ is easily seen to be a bijective and continuous mapping. Since D_n is compact, f_+ is a homeomorphism. Therefore, S^n_+ is a connected set. Similarly, the mapping $f_- : D_n \to S^n_-$ defined by

$$f_-(u) = (u_1, u_2, \ldots, u_n, -\sqrt{1 - \|u\|^2})$$

is a homeomorphism, and so S^n_- is a connected set. Since S^n_+ intersects with S^n_-, $S^n = S^n_+ \cup S^n_-$ is a connected set.

7.5 Connected Subsets of $\mathbb{R}$

In the beginning of this chapter, we mentioned that the intervals in the real line are connected. We now turn to establish this result.

Theorem 7.5.1 A subset of the real line $\mathbb{R}$ is connected if and only if it is an interval. In particular, the real line is connected.

Proof. Let X be a connected subset of $\mathbb{R}$. We show that X is an interval. Suppose on the contrary that X is not an interval. Then there exist real numbers a, b and c such that $a < b < c$, a and c are in X, but b is not in X. Then the sets $X \cap]-\infty, b[$ and $X \cap]b, \infty[$ are non-empty disjoint open sets in X, and X is their union. This shows that X is disconnected, a contradiction. Conversely, we assume that X is an interval, and show that X is connected. Suppose on the contrary that X is not connected. Let (A, B) be a disconnection of X. Since A and B are non-empty, there exist a point x in A and a point z in B. Then $x < z$, or $z < x$. Assume that $x < z$ (In case $z < x$, the proof is similar). Since X is an interval, $[x, z] \subseteq X$. Also each point of $[x, z]$ is either in A or in B. We define $y = \sup\{[x, z] \cap A\}$. Since $x \in [x, z] \cap A$, $x \leq y$. Also, since z is an upper bound of $[x, z] \cap A$, $y \leq z$. Since X is an interval, y is in X. So y is either in A or in B. If y is in A, then $y < z$. Since the interval $]y, z]$ is contained in B, y is a limit point of B, and so $y \in \overline{B}$. This means that $A \cap \overline{B}$ is non-empty, a contradiction since (A, B) is a disconnection of X. If y is in B, then $x < y$, and y is a limit point of A. Therefore, $\overline{A} \cap B$ is non-empty, again a contradiction. This completes the proof.

As a consequence of Theorem 7.4.1 and Theorem 7.5.1, we have the following generalisation of the intermediate value theorem of elementary calculus.

Theorem 7.5.2 The range of a continuous real valued mapping defined on a connected metric space is an interval.

The following Corollary follows from Theorem 6.7.1 and the preceding theorem.

Corollary 7.5.3 The image of a compact and connected metric space under a real valued continuous mapping is a closed and bounded interval.

Since a closed and bounded interval in the real line is compact and connected set, the image under a continuous real valued mapping of a closed and bounded interval will be a closed and bounded interval.

The condition of continuity in Theorem 7.5.2 can not be relaxed, For example, the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \frac{1}{x^3}$ if $x \neq 0$, and $f(0) = 1$ is not continuous at $x = 0$. The range of f, $f(\mathbb{R})$, is not an interval (-1 and 1 are in $f(R)$, but 0 is not in $f(R)$).

Corollary 7.5.4 (Bolzano) Let X be a connected metric space, and $f : X \to \mathbb{R}$ be a continuous mapping. Let a and b any points in $f(X)$ such that $a < b$. Then for any c such

that $a < c < b$, there exists a point x in X such that $f(x) = c$.

Theorem 7.5.5 If α is a non-negative real number, and $n \in \mathbb{Z}^+$, then there exists a unique $x \in [0, \infty[$ such that $x^n = \alpha$.

Proof. The theorem is trivial if $\alpha = 0$. So we assume that $\alpha > 0$. Then there exists a $k \in \mathbb{Z}_+$ such that $k > \alpha$. By considering the continuous mapping $f : [0, \infty[\to [0, \infty[$ given by $f(t) = t^n$ for all t in $[0, \infty[$, we have $f(k) = k^n \geq k > \alpha$, and $f(0) = 0 < \alpha$. Then, by Corollary 7.5.4, there is a point x in $[0, k]$ such that $f(x) = \alpha$. It is easy to see that x is unique.

Theorem 7.5.6 Any polynomial with real coefficients of odd degree has a real root.

Proof. Recall that a polynomial of degree n with real coefficients is a function $p : \mathbb{R} \to \mathbb{R}$ defined by $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$, where $a_0, a_1, \ldots, a_n$ are real numbers, and $a_0 \neq 0$. Dividing, if necessary by -1, we may assume that a_0 is positive . For $x \neq 0$, we write $p(x) = x^n q(x)$, where $q(x) = a_0 + \frac{a_1}{x} + \cdots + \frac{a_n}{x^n}$. Then it follows that $|q(x) - a_0| < \frac{A}{|x|}$, where $|A| = |a_1| + |a_2| + \cdots + |a_n|$, and $|x| > 1$. Therefore, $|q(x) - a_0| < \frac{a_0}{2}$ if $|x| > \max\left\{1, \frac{2A}{a_0}\right\}$. Thus, $q(x) > 0$ if $|x| > \max\left\{1, \frac{2A}{a_0}\right\}$. Let y be a point such that $y > \max\left\{1, \frac{2A}{a_0}\right\}$. Then $q(y)$ and $q(-y)$ are both positive. Since n is odd, $p(y)$ and $p(-y)$ will be of opposite signs. By Corollary 7.5.4, there exists a point α in $[-y, y]$ such that $p(\alpha) = 0$.

EXERCISE

1. Let n be a positive integer, and $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^n$. Show that for each $a \geq 0$, there is exactly one $b \geq 0$ such that $b^n = a$.

2. Prove that a metric space X is connected if and only if every continuous characteristic function on X is constant.

3. Show that the converse of Corollary 7.5.4 is not true, that is, give an example of a metric space X and a mapping $f : X \to \mathbb{R}$ such that for any two points a and b in $f(X)$ such that $a < b$ and a real number c such that $a < c < b$, there exists a point $x \in X$ such that $f(x) = c$, and f is not continuous on X.

4. Let X be a connected metric space, and $f : X \to R$ be a non-constant continuous mapping. Show that X is uncountable.

5. Let (X, d) be an unbounded metric space. Show that for any $x \in X$ and real number $r > 0$, there exists a point y in X such that $d(x, y) = r$.

6. Show that a parabola is not homeomorphic to a hyperbola in $\mathbb{R}^2$.

7. Give an example of a nested sequence $\langle A_n \rangle$ of connected subsets of $\mathbb{R}^3$ such that the intersection $\bigcap_{n=1}^{\infty} A_n$ is not connected.

8. Show that no non-empty open subset of $\mathbb{R}$ is homeomorphic to an open subset of $\mathbb{R}^2$.

7.6 Connectedness of Products

The finite product of connected metric spaces is connected as indicated in the following.

Theorem 7.6.1 Let $X_i,\ i = 1, 2, \ldots, n$ be a metric spaces, and X be their product metric space. Then X is connected if and only if each factor space X_i is connected.

Proof. Let $X_i,\ i = 1, 2, \ldots, n$ be connected metric spaces. We show the connectedness of X by the principle of mathematical induction. The metric space X_1 is connected. Assume by induction that the product of $(X_i, d_i),\ i = 1, 2, \ldots, p,\ p < n$, is connected. Denoting this product by $\prod_{i=1}^{p} X_i$. We show that the product $\prod_{i=1}^{p} X_i \times X_{p+1}$ is connected. To show this, let (x, x_{p+1}) and (x', x'_{p+1}) be any two points in $\prod_{i=1}^{p} X_i \times X_{p+1}$. The set $\{x\} \times X_{p+1}$ is a copy of the connected metric space X_{p+1}, and therefore it is a connected subset of the product $\prod_{i=1}^{p} X_i \times X_{p+1}$, and it contains the points (x, x_{p+1}) and (x, x'_{p+1}). Similarly, the set $\prod_{i=1}^{p} X_i \times \{x'_{p+1}\}$ is a copy of the connected metric space $\prod_{i=1}^{p} X_i$, and is therefore a connected subset of the product $\prod_{i=1}^{p} X_i \times X_{p+1}$, and it contains the points (x, x'_{p+1}) and (x', x'_{p+1}). Since the point (x, x'_{p+1}) is common to $\{x\} \times X_{p+1}$ and $\prod_{i=1}^{p} X_i \times \{x'_{p+1}\}$, by Corollary 7.3.8, their union $S = (\{x\} \times X_{p+1}) \cup \left(\prod_{i=1}^{p} X_i \times \{x'_{p+1}\}\right)$ is connected subset of $\prod_{i}^{p} X_i \times X_{p+1}$, and S contains (x, x_{p+1}) and (x', x'_{p+1}). By Theorem 7.3.14, $\prod_{i=1}^{p} X_i \times X_{p+1}$ is connected. But this product is homeomorphic to the product metric space $\prod_{i=1}^{p+1} X_i$; if $x = (x_1, x_2, \ldots, x_p)$, then f defined by $f(x, x_{p+1}) = (x_1, x_2, \ldots, x_{p+1})$ is a homeomorphism. Therefore, $\prod_{i=1}^{p+1} X_i$ is connected, and the connectedness of the product $\prod_{i=1}^{n} X_i$ follows from the principle of mathematical induction. Conversely, let the product metric space X be connected. Since the projection mappings $\pi_i : X \to X_i$ are continuous surjections, it follows that each X_i is connected.

Theorem 7.6.2 The n-dimensional Euclidean plane $\mathbb{R}^n$ and n-dimensional unitary space $\mathbb{C}^n$ are connected.

Proof. Since $\mathbb{R}^n$ can be regarded as the product of n replicas of the real line $\mathbb{R}$, and $\mathbb{R}$ is connected, $\mathbb{R}^n$ is connected by Theorem 7.6.1. We next show that $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ are essentially same as metric spaces by exhibiting a homeomorphism between $\mathbb{C}^n$ and $\mathbb{R}^{2n}$. Let $z = (z_1, z_2, \ldots, z_n)$ be an arbitrary point of $\mathbb{C}^n$. Let $z_k = a_k + ib_k$ where a_k and b_k are real and imaginary parts of z_k, respectively. We define f by

$$f(z) = (a_1, b_1, a_2, b_2, \ldots, a_n, b_n).$$

Then $f : \mathbb{C}^n \to \mathbb{R}^{2n}$ is a bijective mapping. If the metrics on $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ are denoted by d_1 and d_2, respectively, then it follows that $d_1(z, z') = d_2(f(z), f(z'))$. Hence, both f and f^{-1} are continuous. Therefore, f is a homeomorphism, and so $\mathbb{C}^n$ is connected.

In the following section, we shall introduce maximal connected sets and generalize Theorem 7.6.1.

7.7 Components

If a metric space X is not connected, then we are interested in maximal connected subsets of X known as components.

Definition 7.7.1 A subset A of a metric space X is a component of X if A is a maximal connected subsets of X, that is, A is connected, and there is no connected subset of X which

properly contains A.

Examples 7.7.2

1. Define $X = \{z \in \mathbb{C} : |z| \leq 1\} \cup \{z \in \mathbb{C} : |z-3| < 1\}$, and X be given the subspace metric of $\mathbb{C}$. Then the subset $A = \{z : |z| \leq 1\}$ of X is both open and closed. Then the subset $B = X - A$ is also both open and closed. Thus, X is disconnected. A and B are components of X. Further, these are the only components of X.
2. Let $X = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$ be a metric space with usual metric. Then every component of X is singleton, and each singleton is a component of X. The set $\left\{\frac{1}{n}\right\}$ is open as well as closed for each n. However, the component $\{0\}$ is not an open set.
3. In a discrete metric space, each singleton is a component, and each component is both open and closed.

Theorem 7.7.3 Let (X, d) be a metric space. Then (i) each point in X is contained in a component of X, (ii) each connected subset of X is contained in a component of X, and (iii) distinct components of X are disjoint.

Proof. (i) Let x_0 be a point in X, and $\mathcal{A}$ be the set of connected subsets of X which contains x_0. $\mathcal{A}$ is non-empty since the singleton $\{x_0\}$ is connected set containing x_0. Then the set $B = \cup\{A : A \in \mathcal{A}\}$ is connected by Corollary 7.3.9. The set B is a component of X containing x_0, for if C is another connected set containing B, then $x_0 \in C$, and so C is in $\mathcal{A}$. Thus, B is maximally connected subset of X containing x_0.

(ii) The part (b) is a direct consequence of construction in part (a). Let A be a connected subset of X. Let x be any the point in A. Then x is in a component C of X. Then $A \cap C$ is non-empty, and so $A \cup C$ is connected. Since C is a component, A is contained in C.

(iii) Let A_1 and A_2 be two components of Xsuch that $A_1 \neq A_2$. If $x_0 \in A_1 \cap A_2$, then $A_1 \cup A_2$ is a connected subset of X that properly contains A_1. Thus, A_1 is not a component, a contradiction.

In view of the above theorem, the set of all components of a metric space X gives a partition of X.

Theorem 7.7.4 A component in a metric space is a closed set.

Proof. Let C be a component in a metric space X. Since $\overline{C}$ is connected, and C is a maximal connected subset, it follows that $\overline{C} = C$. Thus, C is a closed set.

The following theorem gives a sufficient condition for a set to be a component.

Theorem 7.7.5 A connected, open and closed subset of a metric space X is a component of X.

Proof. Let A be connected, open and closed subset of X. Since A is connected, by Theorem 7.7.3, there exists a component C of X containing A. If A is properly contained in C, then $C \cap A$ is both open and closed proper subset of A since A is both open and closed. Thus, C is disconnected, a contradiction.

The components in a metric space X can be characterised in another way. We define an equivalence relation on X by setting $x \sim y$ if there is a connected subset of X containing both x and y. This relation is obviously reflexive and symmetric. To see the transitivity assume that $x \sim y$ and $y \sim z$. Then there exist connected subsets A and B such that x and y are in A, and y and z are in B. Then the point y is common to both A and B, and thus $A \cup B$ is connected subset of X, and it contains x and z. Thus, $x \sim z$. The equivalence relation $\sim$ thus gives the partition of X into equivalence classes, and if we denote the equivalence class of x by $[x]$, then we have the following.

Theorem 7.7.6 Let X be a metric space, and x be a point in X. Let C_x be the component of X containing x. Then $C_x = [x]$.

Proof. The easy proof is left to the reader.

Theorem 7.7.7 Let $\{(X_i, d_i)\}$ be a denumerable family of metric spaces, and (X, D) be their product. Then (X, D) is connected if and only if each factor space (X_i, d_i) is connected.

Proof. If X is connected, then each X_i is connected since the projection are continuous surjections. Conversely, assume that each metric space (X_i, d_i) is connected. Let $z = (z_i)$ be a point in X, and C be the component containing z. We prove that $C = X$. Let $x = (x_i)$ be any point in X. Let $\varepsilon > 0$ be a real number. Then there exists an $n_0 \in \mathbb{Z}_+$ such that $\frac{1}{i} < \varepsilon$ for all $i \geq n_0 + 1$ so that $\frac{\overline{d}_i(x_i, y)}{i} < \varepsilon$ for all y in X_i. Now let $Y = \prod_{i=1}^{\infty} A_i$, where $A_i = X_i$ for $i = 1, 2, \ldots, n_0$, and $A_i = \{z_i\}$ for $i \geq n_0 + 1$. Then the subspace Y of (X, D) is homeomorphic to the product $X_1 \times X_2 \times \cdots \times X_n$, and hence it is connected by Theorem 7.6.1. Since $z \in Y$, Y is contained in C. Since the point $t = (t_i)$ where $t_i = x_i$ if $i \leq n_0$, and $t = z_i$ if $i \geq n_0 + 1$ is in Y, and also it is in the open sphere $S(x, \varepsilon)$, it follows that $S(x, \varepsilon)$ intersects with C. Thus, x is in the closure of C. Since C, as a component, is closed, $x \in C$.

Theorem 7.7.8 Let G be an open set in $\mathbb{C}$. Then the components of G are open, and there are countable number of them.

Proof. Let C be a component of G. Let x_0 be a point in C. Since G is open, there exists a real number $\varepsilon > 0$ such that $S(x_0, \varepsilon)$ is contained in G. By Corollary 7.3.8 $S(x_0, \varepsilon) \cup C$ is a connected, and so must be equal to C. Therefore, $S(x_0, \varepsilon)$ is contained in C, and C is an open set. To see that there are only countable number of components of G, let $S = \{a + ib : a \text{ and } b \text{ are rational numbers and } a + ib \text{ is in } G\}$. Then S is countable, and each component of G contains an element of S, so the number of components is countable.

EXERCISE

1. If a metric space (X, d) has only a finite number of components, show that each component is both open and closed.
2. Let G be an open and connected set in a metric space X. Show that G is a component of the complement in X of the boundary of G.
3. If the components of a compact metric space are all open, then show that there are at most finite number of them.
4. Each singleton is a component in the metric space $\mathbb{Q}$ of rational numbers with usual metric.

7.8 Totally Disconnected Metric Spaces

On the opposite extreme of connected metric spaces are totally disconnected metric spaces.

Definition 7.8.1 A metric space X is said to be totally disconnected if for each pair of distinct points x and y in X, there exists a disconnection (A, B) of X such that x is in A, and y is in B. A subset of a metric space X is said to be totally disconnected if it is totally disconnected as a subspace of X.

Examples 7.8.2

1. Every discrete metric space X is totally disconnected. For distinct points x and y in X, the sets $A = \{x\}$ and $B = X - \{x\}$ give the required disconnection of X.
2. A metric space with only one point is totally disconnected; oddly enough, this metric space is also connected.
3. The set of rationals with usual metric is totally disconnected. For any two rational number x and y with $x < y$, there exists an irrational number z between x and y such that the sets $A = \{x \in \mathbb{Q} : x < z\}$ and $B = \{x \in \mathbb{Q} : x > z\}$ give the required disconnection of the set of rationals. Similarly, the set of irrationals with usual metric is totally disconnected.
4. The Cantor set C is totally disconnected.
5. The set $A =]1, 2[\cup]2, 3[$ with usual metric is disconnected, but not totally disconnected. Let x and y be any two distinct points in $]1, 2[$. Suppose that there is a disconnection (B, C) of A such that x is in B, and y is in C. Then $B \cap]1, 2[$ and $C \cap]1, 2[$ are non-empty open sets in $]1, 2[$ whose union is $]1, 2[$. Thus, $]1, 2[$ is disconnected, a contradiction.

Theorem 7.8.3 The components of a totally disconnected metric space X are singletons.

Proof. Let C be a connected subset of X containing the distinct points x and y. Since X is totally disconnected, there exists a disconnection (A, B) in terms of open sets in X such that x is in A and y is in B. Then the sets $A \cap C$ and $B \cap C$ are non-empty, disjoint and open sets in C whose union is C. Thus, C is disconnected, a contradiction. Therefore, any subset of X containing more than one points is disconnected, and the assertion of the theorem follows.

The following theorem gives a sufficient condition for a metric space to be totally disconnected.

Theorem 7.8.4 If a metric space X has an open base of closed sets, then it is totally disconnected.

Proof. Let x and y be any two distinct points in X. Then there exists open set G in X such that G contains x but not y. Let $\mathcal{B}$ be an open base for X such that each member of $\mathcal{B}$ is also a closed set. Then there is a $B \in \mathcal{B}$ such that $x \in B \subseteq G$. Since B is both open and closed, (B, B'), where $B' = X - B$, is a disconnection of X such that $x \in B$ and $y \in B'$.

If X is a compact metric space in the above theorem, then the implication in the above theorem can be reversed.

Theorem 7.8.5 Let X be a compact and totally disconnected metric space. Then X has an open base whose members are also closed in X.

Proof. Let $\mathcal{B}$ be the set of all open and closed sets in X. Let G be an open set, and x be a point in G. If $G = X$, then G is in $\mathcal{B}$, and $x \in G \subseteq G$. So, we assume that $G \neq X$. Since G', the complement of G in X, is closed, and X is compact, G' is compact. Since X is totally disconnected, for each y in G', there exists a disconnection (H_x, H_y), where H_x and H_y are both open and closed in X, of X such that $x \in H_x$ and $y \in H_y$. The collection of H_y's obtained in this way is an open cover of G'. Since G' is compact, there exists a finite subcollection $\{H_{x_1}, H_{x_2}, \ldots, H_{x_n}\}$ which also covers G', but this union does not contain x. Define $H = \bigcup_{i=1}^{n} H_{x_i}$. Then H is both open and closed, and x is not in H. Therefore, H' is also both open and closed, and hence H' is in $\mathcal{B}$. Thus, $x \in H' \subseteq G$.

EXERCISE

1. Is a metric space discrete if its every component is a singleton?
2. Give an example of a totally disconnected metric space in which the closure of an open sphere $S(x, r)$ is the closed sphere $S[x, r]$.
3. Let X be a compact metric space such that the closure of an open sphere $S(x, r)$ is the closed sphere $S[x, r]$. Show that every open sphere in X is connected.
4. Show that the product of two totally disconnected metric spaces is totally disconnected.

7.9 Path Connectedness

The notion of a path in mathematics is abstraction of the real life meaning of the word "path" between any two points. Intuitively, this path consists of connected parts of any zig-zag type.

Definition 7.9.1 Let X be a metric space, and x and y be any two points in X. A path in X from x to y is a continuous mapping $f : [a, b] \to X$ of some closed interval $[a, b]$ in the real line $\mathbb{R}$ into X such that $f(a) = x$ and $f(b) = y$.

Definition 7.9.2 A metric space X is said to be path connected if for every two points x and y in X, there is a path in X from x to y. A subset of X is path connected if it is path connected as a subspace of X.

Examples 7.9.3

1. Any interval in $\mathbb{R}$ is path connected.
2. The following sets in $\mathbb{R}^2$ are path connected.

   (i) a circle (ii) a parabola (iii) an ellipse (iv) union of the lines intersecting at the origin (v) union of any two intersecting parabolas.
3. A hyperbola in $\mathbb{R}^2$ is not path connected.
4. The unit closed sphere $S[0, 1]$ in $\mathbb{R}^n$ centered at origin $0 = (0, 0, \ldots, 0) \in \mathbb{R}^n$ and with radius 1 is path connected. The straight line path $f : [0, 1] \to \mathbb{R}^n$ defined by $f(t) = (1 - t)x + ty$ lies in $S[0, 1]$, for if $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are any points in $S[0, 1]$, then the distance of $f(t)$ from origin, 0 is

$$d(f(t), 0) = \left[\sum_{i=1}^{n} [(1-t)x_i + ty_i]^2\right]^{1/2} \leq (1-t)\left[\sum_{i=1}^{n} x_i^2\right]^{1/2} + t\left[\sum_{i=1}^{n} y_i^2\right]^{1/2} \leq 1.$$

   Similarly, every open sphere and every closed sphere in $\mathbb{R}^n$ can be shown to be path connected.
5. For $n \geq 2$ the subspace $X = \mathbb{R}^n - \{0\}$ of $\mathbb{R}^n$ is path connected. If a and b are any two points in $\mathbb{R}^n$, then the straight line segment joining a and b is a path from a to b in X if the origin does not lie on this segment, otherwise we can choose a point c not on the straight line segment from a to b such that the straight line segments from a to c and from c to b constitute a path from a to b. (The definition of the straight line segment

in $\mathbb{R}^n$ is an obvious generalisation of the definition of the straight line segment in $\mathbb{R}^2$ (see definition 1.14.3).

Theorem 7.9.4 A path connected metric space X is connected.

Proof. Let (A, B) be a disconnection of X. Let $f : [a, b] \to X$ be a path in X. Since f is continuous, $f([a, b])$ is connected subset of X. Then $f([a, b])$ is either contained in A or in B. Therefore, there is no path joining a point in A to a point in B, a contradiction. Therefore, X is connected.

The converse of the above theorem does not hold. A connected metric space may not be path connected as shown in the following.

Example 7.9.5 Define $K = \left\{\frac{1}{n} : n \in \mathbb{Z}_+\right\}$, and consider the subspace $C = ([0, 1] \times \{0\}) \cup (K \times [0, 1]) \cup (\{0\} \times [0, 1])$ of $\mathbb{R}^2$ shown in Fig. 7.1. C is called a combspace.

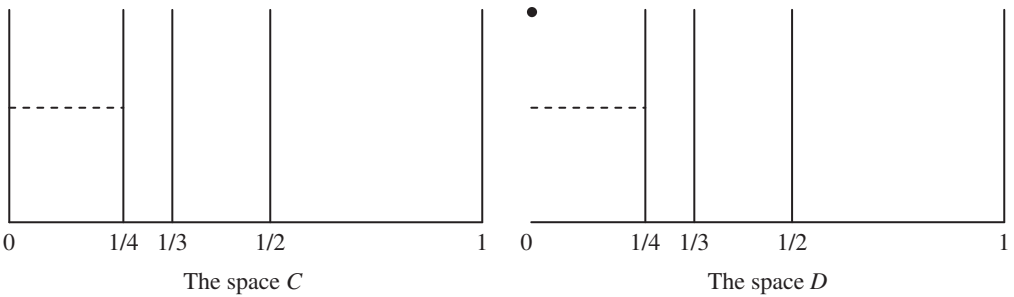


Fig. 7.1

The space C is clearly path connected, and hence connected. The deleted combspace D is obtained from C by deleting the set $\{0\} \times]0, 1[$ shown above. In the metric space D, the set $A = ([0, 1] \times \{0\}) \cup (K \times [0, 1])$ is clearly path connected, and thus connected. The point (0,1) is a limit point of A. Therefore, $\overline{A} = D$, and thus D is connected. The metric space D is not path connected. Let $f : [a, b] \to D$ be a path in D beginning at $p = (0, 1)$. We show that $f^{-1}\{p\})$ is both open and closed in $[a, b]$.

Since f continuous, and the set $\{p\}$ is closed in D, $f^{-1}(\{p\})$ is closed in $[a, b]$. To show that $f^{-1}(\{p\})$ is an open set in $[a, b]$, let x_0 be a point in $f^{-1}(\{p\})$ so that $f(x_0) = p$. Let $r > 0$ be a real number such that the open sphere $S(p, r)$ does not intersect with $x-$axis (see Fig. 7.2). Since f is continuous at x_0, there exists a real number $\delta > 0$ such that $f(S(x_0, \delta))$ is contained in $S(p, r)$. Since the open sphere $S(x_0, \delta)$ is an interval, it is connected. Therefore, the set $f(S(x_0, \delta))$ is a connected subset of D that contains p. Then $f(S(x_0, \delta))$ does not contain any point of D other than p, for assume that $q = (\frac{1}{n}, t_0)$ be a point of D in $S(p, r)$. Choose a real number r' such that $\frac{1}{(n+1)} < r' < \frac{1}{n}$. Then consider the disjoint open subsets $]-\infty, r'[\times\mathbb{R}$ and $]r', \infty[\times\mathbb{R}$ of $\mathbb{R}^2$. Now since $f(S(x_0, \delta))$ is contained in $S(p, r)$, it does not touch x-axis, and the line $x = r$. Therefore, $f(S(x_0, \delta))$ lies in the union $(]-\infty, r'[\times\mathbb{R}) \cup (]r', \infty[\times\mathbb{R})$. Since $f(S(x_0, \delta))$ is connected, and contains p, it is contained in the set $]-\infty, r'[\times\mathbb{R}$, and does not contain the point q of the set $]r', \infty[\times\mathbb{R}$. Therefore, $f(S(x_0, \delta)) = \{p\}$.

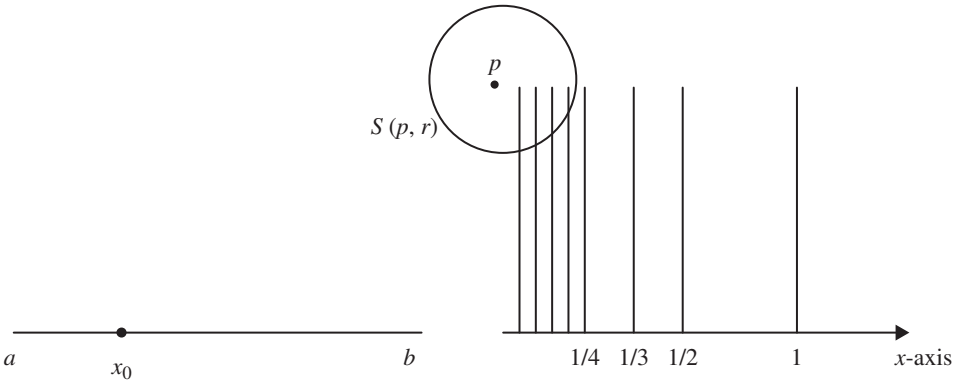


Fig. 7.2

Theorem 7.9.6 A continuous image of a path connected metric space is path connected.

Proof. Let X be a path connected metric space, and Y be a metric space. Let $f : X \to Y$ be a continuous mapping. We show that $f(X)$ is path connected subspace of Y. Let a and b be any two points in $f(X)$. Then, there are points x and y in X such that $f(x) = a$ and $f(y) = b$. Since X is path connected, there is a path $g : [0,1] \to X$ from x to y. Then $g(0) = x$ and $g(1) = y$. Thus, the composition $f \circ g : [0,1] \to Y$ is a path in $f(X)$ from a to b.

The connectedness of open sets in $\mathbb{R}^n$ is equivalent to their path connectedness. The proof of this fact uses the path connectedness of convex sets in $\mathbb{R}^n$. Let us recall the following definition.

Definition 7.9.7 Let V be a vector space over the field of real numbers, and $x, y \in V$. We define $[x, y] = \{(1-t)x + ty \, : \, 0 \leq t \leq 1\}$. The set $[x, y]$ is called a straight line segment in V joining x and y. A subset A of V is said to be convex if for any points x and y in A, the straight line segment $[x, y]$ is contained in A.

We know that $\mathbb{R}^n$ is a vector space over the field of real numbers if we define the addition and scalar multiplication in $\mathbb{R}^n$ by

$$\begin{aligned} x + y &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n), \\ \text{and} \qquad \alpha x &= (\alpha x_1, \alpha x_2, \ldots, \alpha x_n) \end{aligned}$$

for all $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, and $\alpha \in \mathbb{R}$. With algebraic operations defined co-ordinate wise as above, the origin $0 = (0, 0, \ldots, 0)$ is the additive identity in $\mathbb{R}^n$, and $-x = (-x_1, -x_2, \ldots, -x_n)$ is the additive inverse of $x = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$. With its Euclidean metric, it is easy to see that the open spheres in $\mathbb{R}^n$ are convex sets. The mapping $f : [0,1] \to \mathbb{R}^n$ defined by $f(t) = (1-t)x + ty$ for all $t \in [0,1]$ is continuous for any two fixed points x and y in $\mathbb{R}^n$; for this mapping $f[0,1] = [x, y]$. Hence, any convex set in R^n is path connected.

Theorem 7.9.8 Any open and connected subset of $\mathbb{R}^n$ is path connected.

Proof. Let G be an open and connected subset of $\mathbb{R}^n$, and x_0 be a point in G. Let A be a set of all x in G such that there is a path from x to x_0. Since there is a constant path

$\gamma : [0,1] \to \mathbb{R}^n$ given by $\gamma(t) = x_0$ for all t in [0,1], $x_0 \in A$. We show that A is both open and closed in G. Let x in G be a limit point of A. Then there exists a sequence $\langle x_n \rangle$ in A such that $\langle x_n \rangle$ converges to x in $\mathbb{R}^n$. Since x is in G, there exists a real number $\varepsilon > 0$ such that $S(x, \varepsilon)$ is contained in G. By the convergence of $\langle x_n \rangle$ to x, there is an $n_0 \in \mathbb{Z}_+$ such that $x_n \in S(x, \varepsilon)$ for all $n \geq n_0$. Since $x_{n_0} \in A$, there is path from x_{n_0} to x_0. Also since $S(x, \varepsilon)$ is path connected, there is a path from x_{n_0} to x. Thus, there exists a path joining x and x_0. Therefore, x is in A, and A is a closed set. Now let x be any point in A. Then $x \in G$, and so there is a real number $r > 0$ such that $S(x, r) \subseteq G$. Since there is a path from x to x_0, and $S(x, r)$ is path connected, there is a path from x_0 to any point in $S(x, r)$. Thus, $S(x, r) \subseteq A$, and A is open. Hence, $A = G$, otherwise G would be disconnected, a contradiction. Hence G is path connected.

Usually, it is enough to consider paths consisting of straight line segments called polygons in order to decide connectedness of open sets in $\mathbb{R}^n$.

Let a and b be any points in $\mathbb{R}^n$. A polygon from a to b in $\mathbb{R}^n$ is a set $P = \bigcup_{k=1}^{n} [x_{k-1}, x_k]$, where $x_0 = a$ and $x_n = b$, and denote $P = [a, x_1, x_2, \ldots, x_{n-1}, b]$.

Theorem 7.9.9 An open set G in $\mathbb{R}^n$ is connected if and only if for any points x and y in G, there is a polygon from x to y contained in G.

Proof. Suppose on the contrary that an open set G in R^n is not connected. Then $G = A \cup B$, where A and B are both non-empty disjoint open and closed sets in G. Let $a \in A$, and $b \in B$. By hypothesis there is a polygon P from a to b such that $P \subseteq G$. Then there exists a straight line segment $[p, q]$ in P such that $p \in A$ and $q \in B$. Let us define

$$S = \{s \in [0,1] : sq + (1-s)p \in A\},$$

and

$$T = \{s \in [0,1] : sq + (1-s)p \in B\}.$$

Then S and T are disjoint, $S \cup T = [0,1]$, $\quad 0 \in S$, and $1 \in T$. The sets S and T are both open in $[0,1]$, contradicting the connectedness of $[0,1]$. Conversely, let G be an open and connected set in R^n containing two points a and b. Let us define the set

$$A = \{x \in G : \text{there is a polygon } P \subseteq G \text{ from } a \text{ to } x\}.$$

As in the preceding theorem, A is both open and closed in G. Then $A = G$. Thus, there is polygon from a to b.

Corollary 7.9.10 Let G be an open and connected subset of $\mathbb{R}^n$, and a and b be any two points in G. Then there is a polygon P in G from a to b which is made up of line segments parallel to n axes. (Each $\mathbb{R}$ in the Cartesian product $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (n times) is considered as an axis.)

Proof. Let x_0 be an element in G. Define

$$A = \{x \in G : \text{there is a polygon } P \subseteq G \text{ having segment parallel to one of the axes from } x_0 \text{ to } x\}.$$

Then A is both open and closed in G, similarly, as in Theorem 7.9.8 which requires only the following comment : Any point in an open sphere $S(x, \varepsilon)$ can be joined by a polygon contained in $S(x, \varepsilon)$ having the straight line segments parallel to the axes. Let $x = (a_1, a_2, \ldots, a_n)$ and $y = (b_1, b_2, \ldots, b_n) \in S(x, \varepsilon)$. Let

$$x_1 = (b_1, a_2, a_3, \ldots, a_n),\ x_2 = (b_1, b_2, a_3, a_4, \ldots, a_n) \ldots,\ x_{n-1} = (b_1, b_2, \ldots, b_{n-1}, a_n).$$

Then the polygon $P = [x, x_1, x_2, \ldots, x_{n-1}, y]$ consists of straight line segments parallel to the axes, and it is contained in $S(x, \varepsilon)$.

The above theorem is particularly useful in determining the connectedness of open sets in the complex plane $\mathbb{C}$ which for most purposes is the same as the two dimensional Euclidean plane $\mathbb{R}^2$.

Example 7.9.11 If A is a countable subset of $\mathbb{R}^2$, then $\mathbb{R}^2 - A$ is connected, that is, if we remove countable number of points from $\mathbb{R}^2$, the resulting subspace of $\mathbb{R}^2$ remains connected. Let z and w be any two points in $\mathbb{R}^2 - A$. Let l be the right bisector of the line segment joining w and z. Since w and z are not in A, a polygon $[w, u, z]$, where u is a point on l, from w to z if intersects with A, then it will intersect with A in points different from w and z. Suppose, if possible, each polygon $[w, u, z]$, where u is a point on l, intersects with A. Then $\cup\{[w, u, z] : u \in l\} \cap A$ is an uncountable subset of A, a contradiction since A is a countable set. Thus, there is polygon $[w, u, z]$, where u is a point on l (Fig. 7.3), joining w and z which is disjoint from A, that is, contained in $\mathbb{R}^2 - A$. Then $\mathbb{R}^2 - A$ is path connected, and so connected.

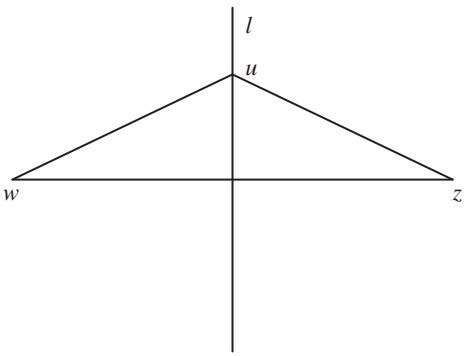


Fig. 7.3

Definition 7.9.12 Let a and b be any two points in a metric space X. A collection $A_1, A_2, \ldots, A_n$ of subsets of X is called a simple chain from a to b if $a \in A_1$, a is not in $A_2, \ldots, A_n$, $b \in A_n$, b is not in $A_1, A_2, \ldots, A_{n-1}$ and $A_i \cap A_j$ is non-empty if and only if $|i - j| \leq 1$.

Fig. 7.4 illustrates a simple chain from a to b

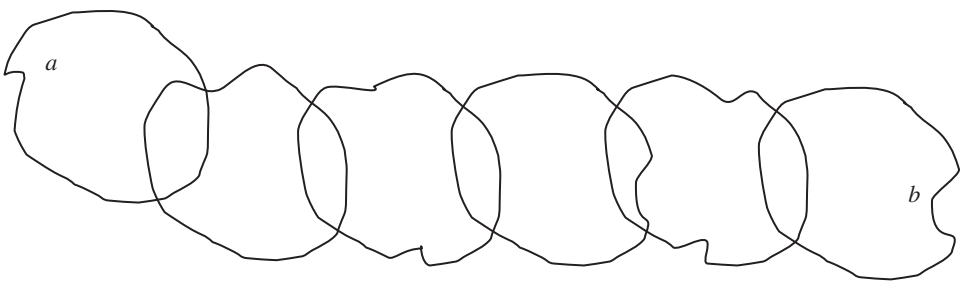


Fig. 7.4

Theorem 7.9.13 Let a and b be two points in a connected metric space X, and $\{G_i\}$ be an open cover of X. Then there exists a simple chain of elements in $\{G_i\}$ from a to b.

Proof. Let Y be the set of all points x in X such that there is a simple chain of elements in $\{G_i\}$ from a to x. Then Y is an open set in X. To see this, let x be a point in Y. Then there exists a simple chain $A_1, A_2, \ldots, A_n$ in $\{G_i\}$ from a to x. If y is any point in A_n,

then there is a simple chain $A_1, A_2, \ldots, A_n$, or $A_1, A_2, \ldots, A_{n-1}$ in the case when y is in $A_{n-1} \cap A_n$, from a to y. Then the whole of A_n is contained in Y. Therefore, Y is a union of open sets. The set Y is also a closed set in X. Let y be a point in $\overline{Y} - Y$. Since $\{G_i\}$ is an open cover of X, there exists a G_{i_0} in $\{G_i\}$ containing y. Since y is a limit point of Y, G_{i_0} contains a point x of Y. Now there is a simple chain $A_1, A_2, \ldots, A_n$ in $\{G_i\}$ from a to x. Therefore, $A_1, A_2, \ldots, A_n, G_{i_0}$ contains a simple chain from a to y. Fig. 7.5 indicates that this simple chain may have less than $n+1$ elements. Thus, y is in Y, a contradiction. Since X is connected, and Y is both open and closed in X, $X = Y$.

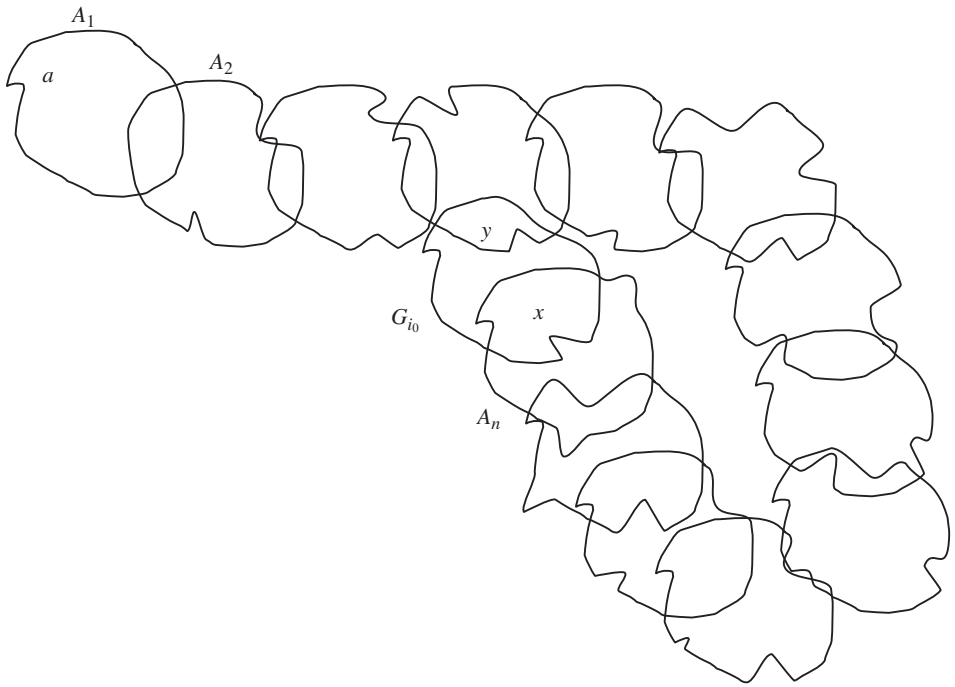


Fig. 7.5

The necessary part of Theorem 7.9.9 is being revisited in the following result.

Theorem 7.9.14 Any two points of an open and connected set G in $\mathbb{R}^n$ can be joined by a polygon in G.

Proof. Let $\{S(x_\alpha, r_\alpha)\}$ be a collection of open spheres that covers G. Such cover exists since the collection of all open sphere in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$. If a and b are any two points in G, then, by Theorem 7.9.13, there is a simple chain $S(x_1, r_1), S(x_2, r_2), \ldots, S(x_n, r_n)$ in $\{S(x_\alpha, r_\alpha)\}$ from a to b . Let y_i be a point in $S(x_i, r_i) \cap S(x_{i+1}, r_{i+1})$ for $i = 1, 2, \ldots, n-1$. Then the polygon $[a, x_1] \cup [x_1, y_1] \cup [y_1, x_2] \cup \ldots \cup [y_{n-1}, x_n] \cup [x_n, b]$ is contained in G.

Example 7.9.15 The unit sphere $S^n = \left\{x = (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : \|x\| = 1\right\}$ is path connected. In order to visualize this example, let us take $n = 2$. Any two points in S^2 lie on a great circle (which gives a path from one point to the other), and this great circle is the intersection with S^2 of a plane passing through the origin.

A 'plane' through the origin in $\mathbb{R}^{n+1}$ is a two dimensional subspace of $\mathbb{R}^{n+1}$. A great

circle on S^n is the intersection of S^n with a plane P passing through the origin. Let $\{u, v\}$ be an orthonormal basis [see Halmos [22]) of the plane P. Then $P \cap S^n$ is given by

$$P \cap S^n = \{x \in S^n : x = (\cos t)u + (\sin t)v \text{ where } t \in [0, 2\pi]\}.$$

Clearly, $x = (\cos t)u + (\sin t)v \in P$ for all $t \in [0, 2\pi]$, and since

$$x \cdot x = \cos^2 t(u \cdot u) + 2\cos t \sin t(u \cdot v) + \sin^2 t(v \cdot v) = 1,$$

$x \in S^n$. Conversely, if $x \in P \cap S^n$, then $x \in P$ and $x \in S^n$. Since $x \in P$, $x = au + bv$ for some $a, b \in \mathbb{R}$. Since $x \in S^n$, $\|x\| = 1$ so that

$$x \cdot x = a^2 + b^2 = 1.$$

Therefore, there exists a $t \in [0, 2\pi]$ such that $a = \cos t$ and $b = \sin t$. This proves our claim for the set $P \cap S^n$. In fact, we have proved that $P \cap S^n$ is the image of $[0, 2\pi]$ under the continuous mapping

$$\gamma_p(t) = (\cos t)u + (\sin t)u \quad \forall\ t \in [0, 2\pi].$$

Note that γ_p is a path that connects any two points on $P \cap S^n$.

Now we show that any distinct points x, y in S^n lie on a great circle. If $y \neq -x$, then x and y are linearly independent, and thus span a two dimensional vector subspace P of $\mathbb{R}^{n+1}$. Then clearly $x, y \in P \cap S^n$, and hence are path connected by the path γ_p. Note that γ_p need not be of the form $\gamma_p(t) = (\cos t)x + (\sin t)y$ unless $x \cdot y = 0$ (To define γ_p that involves only x and y , one needs to recall Gram-Schmidt process [22]). If $y = -x$, since $n \geq 1$, we can find a vector v which is of unit norm, and such that $x \cdot v = 0$. Then $\{x, v\}$ forms an orthonormal basis of a vector subspace P spanned by $\{x, v\}$. Now we define $\gamma_p(t) = (\cos t)x + (\sin t)v$. Then $\gamma_p(0) = x$ and $\gamma_p(\pi) = y$. Thus, x and y are path connected by the path γ_p.

Alternatively, an easier proof is the following. By Example 7.4.6, S^n is the union of two hemispheres S^n_+ and S^n_-, and $S^n_+ \cap S^n_-$ is non-empty. Since the hemispheres are homeomorphic to the convex set D_n which is path connected, the path connectedness of S^n follows.

The similar techniques as were used to define components of a metric space in section 7.7 can be used to define path components. Let X be a metric space. Define a relation $x \sim y$ if there is a path in X from x to y. Then the relation $\sim$ is easily seen to be an equivalence relation on X, and thus it partitions X into equivalence classes. Each equivalence class is called a path component of X. The following theorem summarizes the properties of these path components, and its proof is left as an exercise.

Theorem 7.9.16 Let X be a metric space. (a) Each point of X is in a path component of X, (b) the distinct path components are disjoint, (c) the union of all path components in X is X, (d) each path connected subset of X intersects only one path component of X.

Examples 7.9.17

1. The path components of the subspace $[-1, 0[\cup]0, 1[$ of $\mathbb{R}$ are the sets $[-1, 0[$ and $]0, 1[$.

2. The deleted comb-space D of Example 7.9.5 has two path components $\{(0, 1)\}$ and $(K \times [0, 1]) \cup ([0, 1] \times \{0\})$. If we add to the deleted combspace the points in $\{0\} \times [0, 1]$ having the irrationals as the second co-ordinate, then there will be uncountably many path components; the singleton containing such a point will be a path component.

3. The subspace of rationals in the real line $\mathbb{R}$ will have singeltons only as its path component. Similarly, the path components of the subspace of irrationals are singletons only.

4. The path components of a discrete metric space are singletons only.

EXERCISE

1. Show that the annulus $\{(x_1, x_2) \in \mathbb{R}^2 : 1 \leq \sqrt{x_1^2 + x_2^2} \leq 2\}$ is path connected. Show also that the set $\{(x_1, x_2) \in \mathbb{R}^2 : 1 < \sqrt{x_1^2 + x_2^2} < 2\}$ is path connected.
2. Let A be a connected subset in $\mathbb{R}^n$, and $\varepsilon > 0$. Show that the set $D_\varepsilon(A) = \{x \in R^2 : d(x, A) < \varepsilon\}$ is path connected.
3. Show that the union of the parabolas $A = \{(x, y) \in \mathbb{R}^2 : y^2 = x\}$ and $B = \{(x, y) \in \mathbb{R}^2 : y^2 = -x\}$ is path connected.
4. Show that the subspace $X = \left\{(x, \sin \frac{1}{x}) : x > 0\right\} \cup \{(x, 0) : -1 \leq x \leq 0\}$ of $\mathbb{R}^2$ is connected, but not path connected.
5. Let $f : [0, 1] \to \mathbb{R}^n$ be a path from a point a in the open sphere $S(0, 1)$, where 0 is the origin in $\mathbb{R}^n$, to a point b in $\mathbb{R}^n$ such that $d(0, b) > 1$, where d is the Euclidean metric of R^n. Show that there exists a point $t \in [0, 1]$ such that $d(f(t), 0) = 1$.
6. Show that the unit sphere S^n in $\mathbb{R}^{n+1}$ is a continuous image of the path connected space $\mathbb{R}^{n+1} - \{0\}$. Hint: Consider the mapping $g : \mathbb{R}^{n+1} - \{0\} \to S^n$ defined by $g(x) = \frac{x}{\|x\|}$.
7. If A and B are path connected subsets of a metric space, and if $A \cap B$ is non-empty, prove that $A \cup B$ is path connected.
8. Find a path connected subset of a metric space whose closure is not path connected.
9. Show that the product of two path connected metric spaces is path connected.

7.10 Locally Connected Metric Spaces

We have already seen the concept of locally compact metric spaces. There is a general technique to localise a topological property. Let P be a topological property, for examples compactness, connectedness etc. Then a metric space X has the property P at a point x in X if there is an open set in X containing x and having P (or whose closure has property P). The notion of locally compact metric spaces was introduced using the above approach, but this approach is not satisfactory to localise the notion connectedness. To localise the concept of connectedness, we require that there are arbitrarily small open sets. This is precisely indicated in the following definition.

Definition 7.10.1 A metric space X is said to be locally connected at a point x in X if for every open set U containing x, there is a connected open set V in X such that $x \in V \subseteq U$. The metric space X is said to be locally connected if it is locally connected at every point of X. A subset of a metric space X is said to be locally connected if it is locally connected as a subspace of X.

Examples 7.10.2

1. The real line $\mathbb{R}$ is locally connected, each interval and each ray is locally connected. The set of rational numbers in $\mathbb{R}$ is not locally connected at any point.

2. $\mathbb{R}^n$ and $\mathbb{C}^n$ are locally connected metric spaces.
3. A metric space may be locally connected at all but one points. For instance, consider the graph of the function $y = \sin \frac{1}{x}$, $0 < x \leq 1$, together with the origin (Fig. 7.6).

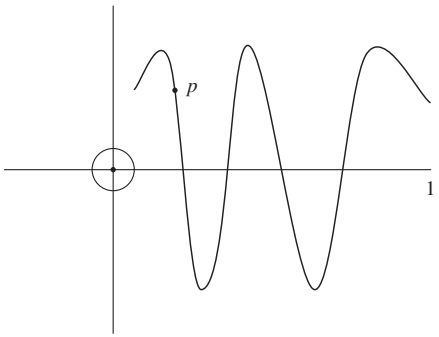


Fig. 7.6

Any open sphere centered at origin such as S shown above defines an open set containing the origin. The only connected set containing the origin and contained in S is the singleton having the origin as its member, but singleton set is not an open set in the space under consideration. Thus, this space is not locally connected at the origin. At any other point in the space such as p, any open set containing p will contain an open arc that contains p, and this open arc is an open and connected set in the space. Thus, the metric space under consideration fails to be locally connected at just one point.

4. The set $[-1, 0[\cup]0, 1[$ with usual metric is locally connected, but not connected.
5. A discrete metric space is locally connected.
6. A finite metric space is locally connected.
7. The deleted comb- space in Example 7.9.5 is connected but not locally connected.
8. Let $A = \left\{(x, \sin \frac{1}{x}) : 0 < x \leq 1\right\}$ and $B = \{(0, y) : -1 \leq y \leq 1\}$ be subsets of $\mathbb{R}^2$. Then $X = A \cup B$ with subspace metric of $\mathbb{R}^2$ is connected, but not locally connected. The set X is usually known as "the topologist sine curve" (Fig. 7.7).

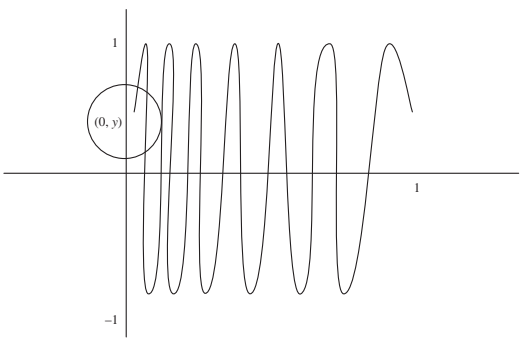


Fig. 7.7 The topologist's sine curve

The set A is the image of the continuous mapping $f :]0,1[\to \mathbb{R}^2$ defined by $f(x) = (x, \sin\frac{1}{x})$. Therefore, A is a connected subset of $\mathbb{R}^2$. Since $X = \overline{A}$, X is connected. To show that X is not locally connected, select any point $(0, y)$ in B. Then any open sphere $S((0, y), \varepsilon)$ intersects both A and B. The only connected set lying within $S((0, y), \varepsilon)$ and containing $(0, y)$ is the line segment of B. But this open segment fails to be neighbourhood of $(0, y)$ since any open sphere centered at $(0, y)$ intersects with A, and thus will not be contained in any segment on B. So, X is not locally connected at every point of B. We note in passing that X is compact since it is a closed and bounded set in $\mathbb{R}^2$.

The topologist's sine curve X is not path connected. To see this, let $f : [0,1] \to X$ be a path such that $f(0) = (0,0)$ and $f(1) = (1, \sin 1)$. We write $f(t) = (f_1(t), f_2(t))$. Since f is continuous, f_1 and f_2 both are continuous. Since B is closed in X, $f^{-1}(B)$ is closed in $[0,1]$. Also $0 \in f^{-1}(B)$, and $f^{-1}(B)$ is bounded. Let t_0 be the supremum of $f^{-1}(B)$. Then $t_0 \in f^{-1}(B)$. Since 1 is in $f^{-1}(A), t_0 < 1$. We show that f_2 is not continuous, and this contradiction establishes our assertion. Let $\delta > 0$ be any real number, such that $t_0 + \delta < 1$. Since $t_0 + \delta$ is not in $f^{-1}(B)$, $f(t_0 + \delta) \in A$. Therefore, $f_1(t_0 + \delta) > 0$. Since $f(t_0) \in B$, $f_1(t_0) = 0$. We assume that $f_2(t_0) \geq 0$. (Obvious modification in the following argument can be made if $f_2(t_0) < 0$). Therefore, there exists an $n \in \mathbb{Z}_+$ such that $f_1(t_0) < \frac{2}{(4n-1)\pi} < f_1(t_0 + \delta)$. By the intermediate value theorem, there exists a t such that $t_0 < t < t_0 + \delta$, and $f_1(t) = \frac{2}{(4n-1)\pi}$. Hence, $f_2(t) = -1$, and $|f_2(t) - f_2(t_0)| > \frac{1}{2}$.

Theorem 7.10.3 Let X be a locally connected metric space. Then (a) each component of X is an open set, and (b) each open subspace of X is locally connected.

Proof. (a) Let C be a component of X, and x be a point in C. Then for each open set U containing x, there is a connected open set V in X such that $x \in V \subseteq U$. Since C is a component, and $C \cap V$ is non-empty, $V \subseteq C$. Therefore, C is an open set in X.

(b) Let Y be a subspace of X such that Y is an open set in X. Let x be any point in Y. Let U be an open set in Y containing x. Since Y is open, U is open in X. Since X is locally connected, there exists a connected open set V in X such that $x \in V \subseteq U$. Since $V \subseteq U \subseteq Y$, V is an open set in Y, and V is connected in Y by Corollary 7.3.5. Therefore, Y is locally connected.

Theorem 7.10.4 A metric space X is locally connected if and only if the components of every open subspace of X are open in X. In particular, the components of X are open.

Proof. First assume that X is locally connected. Let Y be an open subspace of X. Then Y is locally connected by Theorem 7.10.3. Again, by Theorem 7.10.3 (a), each component of Y is open in Y. Since Y is open, each component of Y is open in X. Conversely, assume that the components of an open subspace of X are open in X.Let x be any point in X, and U be an open set in X containing x. Also there exists a component C in the subspace U containing x. By our assumption, C is an open set in X. Thus, C is an open connected set that contains x, and is contained in U.

Theorem 7.10.5 A metric space is locally connected if and only if it has a base of connected open sets.

Proof. The easy proof is left to the reader.

Theorem 7.10.6 The image of a locally connected metric space under an open and continuous mapping is locally connected.

Proof. Let X be a locally connected metric space, and Y be a metric space. Let $f : X \to Y$

be an open and continuous mapping. We show that $f(X)$ is locally connected subspace of Y. Let y be a point in $f(X)$, and H be an open set in $f(X)$ containing y. Then there exists a point x in X such that $f(x) = y$. Since H is open in $f(X)$, there is an open set H^* in Y such that $H^* \cap Y = H$. Since f is continuous, $f^{-1}(H^*)$ is open in X. Also $f^{-1}(H^*) = f^{-1}(H)$. Since X is locally connected, there exists a connected open set G in X such that $x \in G \subseteq f^{-1}(H)$. Then $y \in f(G) \subseteq H$. Since f is open, $f(G)$ is open. Also $f(G)$ is connected, and the proof is complete.

The following result is similar to the above theorem, and the proof is left as an exercise.

Theorem 7.10.7 The image of a locally connected metric space under a closed and continuous mapping is locally connected.

Combining Corollary 6.7.7 and Theorem 7.10.7, we obtain the following.

Theorem 7.10.8 The image of a compact and locally connected metric space under a continuous mapping is a compact and locally connected metric space.

EXERCISE

1. In the Euclidean plane $\mathbb{R}^2$, let

   $A = \{(x_1, x_2) \in \mathbb{R}^2 : x$ is irrational and $0 \leq y \leq 1$, or x is rational and $-1 \leq y < 0\}$

   (a) Show that A is connected and not locally connected.

   (b) Let $f : [0, 1] \to \mathbb{R}$ and $g : [0, 1] \to \mathbb{R}^2$ be continuous mappings, and $h : [0, 1] \to \mathbb{R}^2$ be defined by $h(x) = (f(x), g(x))$. Show that h is continuous, and that f is constant.

2. Prove that the product of two metric spaces is locally connected if and only if each factor space is locally connected.

3. Let $X = \{(x, 0) : x$ is a rational number in $[0, 1]\}$, $p = (0, 1) \in \mathbb{R}^2$, and T be the set of straight line segments $[p, q]$ such that $q \in X$.

   (a) Show that T is path connected, but is locally connected only at the point p.

   (b) Find a subset of $\mathbb{R}$ that is path connected, but is locally connected at none of its points.

4. Show that local connectedness is preserved by a homeomorphism, but need not be preserved by a continuous mapping.

5. A metric space X is said to be locally path connected if for each point x in X and each neighbourhood U of x, there is a path connected neighbourhood V of x which is contained in U. Give an example of a locally path connected metric space. Prove that a metric space which is connected and locally path connected is path connected. Is the converse true?

7.11 Non-homeomorphic Metric Spaces

We have already proved that any two non-degenerate open intervals with usual metrics are homeomorphic to each other. However, a non-degenerate open interval $]a, b[$ is not homeomorphic to a non-degenerate closed interval $[c, d]$. This can be easily seen. Let there be a homeomorphism from $[c, d]$ onto $]a, b[$. Since $[c, d]$ is compact, and compactness is preserved under homeomorphisms, $]a, b[$ will be compact which, of course, is not the case. Next we show that a non-degenerate open interval $]a, b[$ is not homeomorphic to the semi-closed interval $]c, d]$. As both $]a, b[$ and $]c, d]$ are not compact, the above argument shall not be applicable. However, connectedness can be used to prove the assertion. First we note the following property of homeomorphisms: Let (X, d_1) and (Y, d_2) be metric spaces, and $f : X \to Y$ be a homeomorphism. Let a subset A of X is removed from X, and $f(A)$ is removed from Y. Then $X - A$ and $Y - f(A)$ are homeomorphic to each other. Just consider the restrictions of f and f^{-1} to $X - A$ and $Y - f(A)$, respectively. Now let us take $X =]c, d]$, and $Y =]a, b[$ with usual metrics. If we remove the set $\{d\}$ from X, then $X - \{d\}$ is an open interval, and so it is connected, while if we remove $\{f(d)\}$ from Y, the resulting set $]a, f(d)[\cup]f(d), b[$ is disconnected. Thus, if we assume that X and Y are homeomorphic, then $X - \{d\}$ and $Y - \{f(d)\}$ will be homeomorphic, and so $]a, f(d)[\cup]f(d), b[$ will be connected, a contradiction. Similarly, the connectedness argument can be applied to show that the non-degenerate interval $]a, b[$ and $[c, d]$ are not homeomorphic to each other. From the interval $[c, d]$, the removal of the points c and d will leave it connected, while the removal of any two points from $]a, b[$ will make the resulting set disconnected. The connectedness argument can also be used to prove that $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic to each other. The removal of one point from $\mathbb{R}$ will make it disconnected whereas $\mathbb{R}^2$ remains connected after the removal of a single point; infact $\mathbb{R}^2$ remains connected when a countable number of points are removed from it (Example 7.9.11). We can extend our problem and ask: are $\mathbb{R}^m$ and $\mathbb{R}^n$ $(m \neq n)$ homeomorphic?

Chapter 8

Approximations

8.1 The Weierstrass Approximation Theorem

Recall that a polynomial $p : [a, b] \to \mathbb{R}$ is a function defined by $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, where the coefficients $a_0, a_1, \ldots, a_n$ are real numbers. Usually we say that p is a polynomial with real coefficients defined on $[a, b]$. In elementary calculus, it is shown that such a polynomial is continuous. Also, by Theorem 4.7.2, it follows that the limit of uniformly convergent sequence of polynomials is a continuous real valued function. The famous Weierstrass theorem says that the converse of this theorem is also true, that is, any continuous real valued function is the limit of a uniformly convergent sequence of polynomials. The proof given here uses Bernstein polynomials in which we show that any continuous real valued function is the limit of a uniformly convergent sequence of Bernstein polynomials. For another proof, we refer the interested reader to Rudin [42].

Let f be a real valued function defined in the closed interval $[0, 1]$. The Bernstein polynomial $B_n : [0, 1] \to \mathbb{R}$ associated with f is the function defined by

$$B_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(\tfrac{k}{n})$$

where $\binom{n}{k}$ is the binomial coefficient defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

There are several identities that will be needed later on. By the Binomial theorem, the following equality holds

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = [x + (1-x)]^n = 1. \tag{8.1}$$

Differentiating equation (8.1) with respect to x, we get

$$\sum_{k=0}^{n} \binom{n}{k} \left[kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}\right]$$
$$= \sum_{k=0}^{n} \binom{n}{k} x^{k-1}(1-x)^{n-k-1}(k-nx) = 0.$$

Now multiplication of the above equality by $x(1-x)$ gives

$$\sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k}(k-nx) = 0. \tag{8.2}$$

Differentiating equation (8.2) with respect to x gives the equality

$$\sum_{k=0}^{n} \binom{n}{k} [-nx^k(1-x)^{n-k} + x^{k-1}(1-x)^{n-k-1}(k-nx)^2] = 0.$$

The equality (8.1) will reduce the last equality to

$$\sum_{k=0}^{n} \binom{n}{k} x^{k-1}(1-x)^{n-k-1}(k-nx)^2 = n.$$

By multiplying $\frac{x(1-x)}{n^2}$ in the above equality, we get

$$\sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k}(x-\frac{k}{n})^2 = \frac{x(1-x)}{n}. \tag{8.3}$$

Theorem 8.1.1 (Weierstrass) Let f be a continuous real valued function defined on the closed interval $[a,b]$, and $\epsilon > 0$ be a real number. Then there exists a polynomial p with real coefficients defined on $[a,b]$ such that $|f(x) - p(x)| < \epsilon$ for all x in $[a,b]$.

Proof. If $a = b$, then the theorem follows immediately by taking p to be the constant polynomial defined by $p(x) = f(a)$. So we assume that $a < b$. Also observe that the function g_1 defined by $g_1(x') = a + (b-a)x' = x$ is a continuous function of $[0,1]$ onto $[a,b]$. Therefore, the composition $g = f \circ g_1$ is continuous real valued function defined on $[0,1]$. If our theorem is proved for the case in which $a = 0$ and $b = 1$, then there exists a polynomial p' with real coefficients defined on $[0,1]$ such that $|g(x') - p'(x')| < \epsilon$ for all x' in $[0,1]$, then in terms of x this inequality becomes $|f(x) - p'\left(\frac{x-a}{b-a}\right)| < \epsilon$. Now taking

$$p'\left(\frac{x-a}{b-a}\right) = p(x)$$

we get the required inequality $|f(x) - p(x)| < \epsilon$ for all x in $[a,b]$. Thus, we prove the theorem in the case $a = 0$ and $b = 1$.

Let $B_n(x)$ be the Bernstein polynomial associated with f. Then

$$f(x) - B_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k} \left[f(x) - f\left(\tfrac{k}{n}\right)\right].$$

Then

$$|f(x) - B_n(x)| \leq \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left| f(x) - f\left(\tfrac{k}{n}\right) \right|. \tag{8.4}$$

Since f is uniformly continuous on $[0,1]$, there exists a real number $\delta > 0$ such that

$$\left| x - \frac{k}{n} \right| < \delta \quad \text{implies that} \quad \left| f(x) - f\left(\tfrac{k}{n}\right) \right| < \frac{\epsilon}{2}.$$

Let α be the sum of terms in the right of the inequality (8.4) having $\left|x - \frac{k}{n}\right| < \delta$, and β be the sum of other terms, where x is assumed to be arbitrary, but fixed point in [0,1]. Then it follows that $\alpha < \frac{\epsilon}{2}$. We show that if n is sufficiently large, then β is also less than $\frac{\epsilon}{2}$ independently of x. Since f is bounded, there exists a real k such that $|f(x)| \leq k$ for all x in $[0,1]$. Then

$$\beta \leq 2k \sum \binom{n}{k} x^k (1-x)^{n-k} = 2k\gamma$$

where $\gamma = \sum \binom{n}{k} x^k (1-x)^{n-k}$, and, by the definition of β the summation in γ is taken over all k such that $|x - \frac{k}{n}| \geq \delta$. The identity (8.3) just before the statement of the theorem then gives

$$\frac{x(1-x)}{n} \geq \delta^2 \gamma.$$

Therefore, $\beta \leq \frac{2kx(1-x)}{n\delta^2} \leq \frac{k}{2n\delta^2}$ since the maximum value of $x(1-x)$ in $[0,1]$ is $\frac{1}{4}$. If we now choose positive integer $n > \frac{k}{\epsilon\delta^2}$, then $\beta < \frac{\epsilon}{2}$ for all x in $[0,1]$, and the proof is complete.

The Weierstrass theorem in fact proves the following assertion.

Corollary 8.1.2 The set of polynomials with real coefficients is dense in $\mathcal{C}[a,b]$.

Remark. The expression $\binom{n}{k} x^k (1-x)^{n-k}$ in the definition of a Bernstein polynomial is the binomial probability of occurrence of k successes in n trials of an experiment in which the probability of successes in a trial is x, and that of failure is $1-x$. The expected number of successes is nx. If the number of successes k is close to nx, then one is more likely to obtain k successes in n trials. If k is far away from nx, then k successes are less likely to be obtained in n trials. Thus, α is the sum of terms in which k refers to the number of successes which is "more probable" in n trials, while k in any term in the sum β means that the occurrence of k success is "less probable" in n trials.

8.2 Algebra of Continuous Real Valued Mappings

We have already seen in Section 5.11 that the metric space $\mathcal{C}_b(X, \mathbb{R})$ (or $\mathcal{C}_b(X, \mathbb{C})$) of continuous and bounded real (respectively complex) valued mappings defined on a metric space X is a complete normed vector space with respect to pointwise addition, scalar multiplication and sup norm. There will be an occasion in Section-8.3 where it is needed that a third operation called the product of vectors be introduced in these spaces. To take care of such situations, a third operation is introduced in vector spaces.

Definition 8.2.1 A vector space E over a field K is called an algebra if there is a mapping $(x, y) \to xy$ from $E \times E$ into E called the product satisfying the following properties:

1. $x(yz) = (xy)z$,

2. $x(y+z) = xy + xz$ and $(x+y)z = xz + yz$, and

3. $\alpha(xy) = (\alpha x)y = \alpha(xy)$ for every scalar α in K.

4. E is called real or complex algebra according as K is real or a complex field.

   An algebra E is said to be commutative if $xy = yz$

5. An algebra E is said to be with identity if there exists an element 1 in E such that $1x = x1 = x$ for every x in E.

There is a unique identity element in an algebra if it exists.

The spaces $\mathcal{C}_b(X, \mathbb{R})$ and $\mathcal{C}_b(X, \mathbb{C})$ are commutative algebras with identity if we define the product pointwise: $(fg)(x) = f(x)g(x)$, and the identity element is the constant mapping $1(x) = 1$ for all x in X as the following lemma indicates. (Note that the product of mappings is not the same as the composition of mappings.)

Lemma 8.2.2 Let (X, d) be a metric space. Then $\mathcal{C}_b(X, \mathbb{R})$ and $\mathcal{C}_b(X, \mathbb{C})$ are closed under multiplications, that is, $fg \in \mathcal{C}_b(X, \mathbb{R})$ (respectively $\mathcal{C}_b, (X, \mathbb{C})$) for all f and g in $\mathcal{C}_b(X, \mathbb{R})$ (respectively $\mathcal{C}_b(X, \mathbb{C})$). Moreover, if $(f \vee g)(x) = \max\{f(x), g(x)\}$ and $(f \wedge g)(x) = \min\{f(x), g(x)\}$, then $f \vee g$, $f \wedge g \in \mathcal{C}_b(X, \mathbb{R})$ for all f and g in $\mathcal{C}_b(X, \mathbb{R})$.

Proof. Let x_0 be a point in X, and $\epsilon > 0$ be a real number. We choose a real number $\epsilon_1 > 0$ such that $\epsilon_1 (|f(x_0)| + |g(x_0)|) + \epsilon_1^2 < \epsilon$. Since f and g are continuous at x_0, there exist real numbers $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$d(x, x_0) < \delta_1 \quad \text{implies that} \quad |f(x) - f(x_0)| < \epsilon_1$$

and

$$d(x, x_0) < \delta_2 \quad \text{implies that} \quad |g(x) - g(x_0)| < \epsilon_1.$$

Then $d(x, x_0) < \min\{\delta_1, \delta_2\}$ implies the following:

$$\begin{aligned} |(fg)(x) - (fg)(x_0)| &\leq |f(x)g(x) - f(x)g(x_0)| + |f(x)g(x_0) - f(x_0)g(x_0)| \\ &< \epsilon_1(|f(x)| + |g(x_0)|) \\ &\leq \epsilon_1|f(x) - f(x_0)| + \epsilon_1(|f(x_0)| + |g(x_0)|) \\ &< \epsilon_1^2 + \epsilon_1(|f(x_0)| + |g(x_0)) < \epsilon. \end{aligned}$$

This shows that fg is continuous. Since f and g are bounded, fg is bounded. Thus, $fg = \mathcal{C}_b(X, \mathbb{R})$ (respectively $\mathcal{C}_b(X, \mathbb{C})$).

With symbols as defined above, we can now state the following.

Theorem 8.2.3 The real vector space $\mathcal{C}_b(X, \mathbb{R})$ is a commutative algebra with identity with respect to the pointwise multiplication, and $\|fg\| \leq \|f\| \, \|g\|$, and $\|1\| = 1$. Furthermore, $\mathcal{C}_b(X, \mathbb{R})$ is a lattice with respect to the order

$$f \leq g \text{ if } f(x) \leq g(x) \text{ for all } x \text{ in } X$$

in which $f \vee g$ and $f \wedge g$ are join and meet of f and g, respectively.

Proof. From Lemma 8.2.2, it follows that $\mathcal{C}_b(X, \mathbb{R})$ is a commutative algebra with identity, and that it is a lattice with respect to the given partial order. The fact that $\|1\| = 1$ is obvious. Also

$$\|fg\| = \sup\{|f(x)g(x)|\} \leq \sup\{|f(x)|\} \sup\{|g(x)|\} = \|f\| \, \|g\|.$$

A moment of thought will convince the reader of the following parallel result for $\mathcal{C}_b(X, \mathbb{C})$.

Theorem 8.2.4 The complex linear space $\mathcal{C}_b(X, \mathbb{C})$ is a commutative algebra with identity with respect to the pointwise multiplication in which $\|fg\| \leq \|f\| \, \|g\|$, and $\|1\| = 1$. Moreover, if $\bar{f}$ is defined by $\bar{f}(z) = \overline{f(z)}$, then the mapping $f \to \bar{f}$ from $\mathcal{C}_b(X, \mathbb{C})$ into $\mathcal{C}_b(X, \mathbb{C})$ satisfies: $\overline{(f+g)} = \bar{f} + \bar{g}$, $\overline{(\alpha f)} = \bar{\alpha}\bar{f}$, $\overline{(fg)} = \bar{f}\bar{g}$, $\bar{\bar{f}} = f$, and $\|\bar{f}\| = \|f\|$. The mapping $\bar{f}$ is called the complex conjugate of f.

Definition 8.2.5 A subalgebra A of an algebra E is a vector subspace of E such that $xy \in A$ for all x and y in A.

A subalgebra of an algebra is an algebra in its own right.

From the definition of a subalgebra, it follows that the intersection of a family of subalgebras of an algebra E is also a subalgebra of E. If A is a non-empty subset of E, then the intersection of all the subalgebras of E containing A is also subalgebra of E containing A, and this subalgebra is said to be generated by A.

In the remainder of this section, we establish the fact that the closure of a subalgebra of $\mathcal{C}_b, (X, \mathbb{R})$ (respectively, $\mathcal{C}_b(X, \mathbb{C})$) is a subalgebra of $\mathcal{C}_b(X, \mathbb{R})$(respectively, $\mathcal{C}_b(X, \mathbb{C})$).

Theorem 8.2.6 If F is a vector subspace of a normed space E, its closure $\bar{F}$ in E is a vector subspace.

Proof. By assumption, the mapping $(x, y) \to x + y$ of $E \times E$ into E maps $F \times F$ into F, hence maps $\overline{F \times F}$ into $\bar{F}$ by Theorems 4.2.10 and 4.8.13. Since $\overline{F \times F} = \bar{F} \times \bar{F}$ by Theorem 2.15.5, the relations $x \in \bar{F}$, $y \in \bar{F}$ imply $x + y \in \bar{F}$. Using the continuity of $(\lambda, x) \to \lambda x$ we similarly show that $x \in \bar{F}$ implies $\lambda x \in \bar{F}$ for any scalar λ.

Let $E_1, E_2, \ldots, E_n$, F be vector spaces. A mapping $f : E_1 \times E_2 \times \cdots \times E_n \to F$ is called multilinear (or n-linear) if for each fixed value of $n - 1$ arguments, it depends linearly on the remaining argument, that is, let j, $1 \leq j \leq n$, be a fixed number and for $i \neq j$, $z_i \in E_i$ be fixed vectors then for x_j, $y_j \in E_j$,

$$f(z_1, z_2, \ldots, z_{j-1}, \alpha x_j + \beta y_j, z_{j+1}, \ldots, z_n)$$
$$= \alpha f(z_1, \ldots, z_{j-1}, x_j, z_{j+1}, \ldots, z_n) + \beta f(z_1, \ldots, z_{j-1}, y_j, z_{j+1}, \ldots, z_n).$$

Theorem 8.2.7 Let $E_1, \ldots, E_n, F$ be normed spaces, and $f : E_1 \times \cdots \times E_n \to F$ be a multilinear mapping. Then f is continuous if and only if there exists a real number $k > 0$ such that

$$\|f(x_1, \ldots, x_n)\| \leq k\|x_1\| \, \|x_2\| \, \ldots \|x_n\|.$$

Proof. Let us take $n = 2$. We first prove that f is continuous at $(c_1, c_2) \in E_1 \times E_2$. We write $f(x_1, x_2) - f(c_1, c_2) = f(x_1 - c_1, x_2) + f(c_1, x_2 - c_2)$. Hence, $\|f(x_1, x_2) - f(c_1, c_2)\| \leq k\,(\|x_1 - c_1\| \, \|x_2\| + \|c_1\| \, \|x_2 - c_2\|)$. For any real number $\delta > 0$ such that $\delta < 1$, suppose that $\|x_1 - c_1\| \leq \delta$, $\|x_2 - c_2\| \leq \delta$, hence $\|x_2\| \leq \|c_2\| + 1$. We therefore have

$$\|f(x_1, x_2) - f(c_1, c_2)\| \leq k\,(\|c_1\| + \|c_2\| + 1)\,\delta$$

which is arbitrarily small with δ.

Conversely, let f be continuous. Then f is continuous at $(0,0)$. Then there exists an open sphere $S = \{(x_1x_2) : \sup(\|x_1\|, \|x_2\|) < r\}$ in $E_1 \times E_2$ such that the relation $(x_1, x_2) \in S$ implies that $\|f(x_1, x_2)\| \leq 1$. Let now (x_1, x_2) be arbitrary. Suppose first that $x_1 \neq 0$ and $x_2 \neq 0$; then if $z_1 = \frac{rx_1}{\|x_1\|}$, and $z_2 = \frac{rx_2}{\|x_2\|}$, then $\|z_1\| = \|z_2\| = r$, and therefore, $\|f(z_1, z_2)\| \leq 1$. But $f(z_1, z_2) = \frac{r^2 f(x_1,x_2)}{\|x_1\|x_2\|}$, so that $\|f(x_1, x_2)\| \leq k\|x_1\|\|x_2\|$ with $k = \frac{1}{r^2}$. If $x_1 = 0$, or $x_2 = 0$, $f(x_1, x_2) = 0$. Hence, the preceding inequality still holds.

By Theorems 8.2.3, 8.2.4 and 8.2.7, it follows that the mapping $(f, g) \to fg$ is continuous. Now let A be a subalgebra of $\mathcal{C}(X, \mathbb{R})$ (respectively, $\mathcal{C}(X, \mathbb{C})$). Then a proof similar to the one in Theorem 8.2.6 shows that fg is in the closure $\bar{A}$ of A for all $f, g \in \bar{A}$. Thus, we have the following theorem.

Theorem 8.2.8 The closure of a subalgebra A of $\mathcal{C}(X, \mathbb{R})$ (respectively, $\mathcal{C}(X, \mathbb{C})$) is a subalgebra of $\mathcal{C}(X, \mathbb{R})$(respectively, $\mathbb{C}(X, \mathbb{C})$).

8.3 Stone-Weierstrass Approximation Theorem

The Weierstrass theorem considers the metric space $\mathcal{C}[a, b]$ of continuous real valued functions defined on the compact metric space $[a, b]$. We wish to replace the compact metric space $[a, b]$ by an arbitrary compact metric space X, and make a similar statement about the metric space $\mathcal{C}(X, \mathbb{R})$. In this project, we need to specify the meaning of a polynomial defined on X. This entails that we take a closer look at what a polynomial with real coefficient defined on [a,b] is.

A polynomial $p(x) = a_0 + a_1x + \cdots + a_nx^n$ with real coefficients defined on $[a, b]$ can be built from two functions : a constant function 1 defined by $1(x) = 1$ for all x, and the identity function x define by $x(x') = x'$ for all x', using the three operations of pointwise addition, multiplication of real numbers and product: $(f + g)(x) = f(x) + g(x)$, $(af)(x) = af(x)$, $(fg)(x) = f(x)g(x)$. The set P of all polynomials with real coefficients defined on $[a, b]$ is identical with the set of all functions which can be built from the functions 1 and x using the above three operations. It follows from Theorem 8.2.3 that $\mathcal{C}[a, b]$ is a commutative algebra with identity 1 with respect to the pointwise addition, scalar multiplication and the product, and the set P with respect to these three operations is a subalgebra of $\mathcal{C}[a, b]$. Moreover, any subalgebra of $\mathcal{C}[a, b]$ that contains the functions 1 and x will contain P also. Hence, P is the intersection of all the subalgebras of $\mathcal{C}[a, b]$ that contain the set $\{1, x\}$. Therefore, P is the subalgebra of $\mathcal{C}[a, b]$ generated by the set $\{1, x\}$.

Now the Weierstrass approximation theorem can equivalently be stated as follows.

Theorem 8.3.1 A subalgebra of $\mathcal{C}[a, b]$ that contains the set $\{1, x\}$ is dense in $\mathcal{C}[a, b]$.

There is another property of the subalgebra P of $\mathcal{C}[a, b]$ which is crucial for the generalisation of the above theorem. If x and y are distinct points in $[a, b]$, then there exists a polynomial p in P such that $p(x) \neq p(y)$; just take a linear polynomial.

Definition 8.3.2 Let X be a metric space, and $\mathbb{B}(X, \mathbb{R})$ be the set of all bounded real valued mappings defined on X. A subset A of $\mathbb{B}(X, \mathbb{R})$ separates points of X if for any pair of distinct points x and y in X, there exists an element f in A such that $f(x) \neq f(y)$.

Now we are in a position to prove the following generalisation of the Weierstrass approximation theorem.

Theorem 8.3.3 (Stone-Weierstrass) Let X be a compact metric space. If A is a subalgebra of $\mathcal{C}(X, R)$ that contains constant mappings, and separates points of X, then A is dense in $\mathcal{C}(X, \mathbb{R})$.

Proof. The proof is divided in several steps.
Step-1. We define a sequence $\langle p_n \rangle$ of polynomials with real coefficients defined on [0,1] inductively as follows:

$$\begin{aligned} p_1 &= 0, \\ p_{n+1}(t) &= p_n(t) + \frac{(t - p_n^2(t))}{2} \text{ for all } n \geq 1, \text{ and } t \in [0,1]. \end{aligned} \tag{8.5}$$

By induction, we show that $p_n(t) \leq \sqrt{t}$ for all $t \in [0,1]$. Assuming $p_n(t) \leq \sqrt{t}$, we have

$$\begin{aligned} \sqrt{t} - p_{n+1}(t) &= \sqrt{t} - p_n(t) - \frac{(t - p_n^2(t))}{2} \\ &= \Big(\sqrt{t} - p_n(t)\Big)\left(1 - \frac{\big(\sqrt{t} + p_n(t)\big)}{2}\right) \geq 0. \end{aligned}$$

Therefore, $p_{n+1}(t) \leq \sqrt{t}$ for all $t \in [0,1]$, and so $p_n(t) \leq \sqrt{t}$ for all $t \in [0,1]$ and for all $n \in \mathbb{Z}_+$ by induction. Again, by induction on n, we easily show that $p_n(t) \geq 0$ for all $t \in [0,1]$ and $n \in \mathbb{Z}_+$. A third application of induction will show that $p_n \leq p_{n+1}$ for all $n \in \mathbb{Z}_+$. Thus, $\langle p_n \rangle$ is a monotonically increasing sequence of polynomials. Therefore, for each $t \in [0,1]$, the sequence $\langle p_n(t) \rangle$ is monotonically increasing and bounded, and so it is convergent. Letting n to tend to infinity in (8.5), it follows that the limit of $\langle p_n(t) \rangle$ is $v(t)$ which satisfies the equation $t - v^2(t) = 0$, that is, $v(t) = \sqrt{t}$ for all $t \in [0,1]$. Since v is continuous, and $\langle p_n \rangle$ is monotonically increasing, by Theorem 6.7.15, it follows that $\langle p_n \rangle$ converges uniformly to v.

Step-2. Here we show that if $f \in A$, then $|f| \in \bar{A}$, the closure of A in $\mathcal{C}(X, \mathbb{R})$. Now, by Theorem 6.7.10, the supremum $a = \sup\{|f(t)| : t \in X\}$ exists. Therefore, by step 1, the sequence $\langle p_n(\frac{f^2}{a^2}) \rangle$, which is in A by the definition of an algebra, converges uniformly to $\sqrt{\frac{f^2}{a^2}} = \frac{|f|}{a}$. Hence, $|f|$ is in $\bar{A}$.

Step-3. Now we prove that for any f, g in $\bar{A}$, $\inf\{f, g\}$ and $\sup\{f, g\}$ are in $\bar{A}$. Since $\inf\{f, g\} = \frac{f+g-|f-g|}{2}$ and $\sup\{f, g\} = \frac{f+g+|f-g|}{2}$, the assertion follows from Theorem 8.2.8 and step 2.

Step-4. Here we prove that for any pair of distinct points x and y in X, and any pair of real numbers α and β there is a mapping f in $\bar{A}$ such that $f(x) = \alpha$, and $f(y) = \beta$. Since A separates points, there exists a mapping g in A such that $g(x) \neq g(y)$. As A contains constant mappings, the mapping f defined by

$$f(z) = \frac{\alpha(g(z) - g(y))}{g(x) - g(y)} + \frac{\beta(g(z) - g(x))}{g(y) - g(x)}$$

has the required properties.

Step 5. In this step we prove that for any mapping f in $\mathcal{C}(X,\mathbb{R})$, any point x in X and any real number $\epsilon > 0$, there is a mapping g in $\bar{A}$ such that $g(x) = f(x)$, and $g(y) \leq f(y) + \epsilon$ for all y in X. To see this assertion, let z be a point in X. Then there exists a mapping h_z in $\bar{A}$ such that $h_z(x) = f(x)$, and $h_z(z) \leq f(z) + \frac{\epsilon}{2}$; the existence of such a mapping is obvious if $z = x$ (for instance, take the constant mapping $h_z(y) = f(z)$ for all y in X). If $z \neq x$, then the existence of h_z is assured by step 4. Since f and h_z are continuous at z, there exists a neigbourhood U_z of z such that for all y in U_z

$$|h_z(y) - h_z(z)| < \frac{\epsilon}{4} \quad \text{and} \quad |f(y) - f(z)| < \frac{\epsilon}{4}.$$

Then it follows that $h_z(y) \leq f(y) + \epsilon$ for all y in U_z. Since X is compact, it can be covered by a finite number of such neighbourhoods U_z, say,

$$U_{z_1}, U_{z_2}, \ldots, U_{z_n}.$$

By step 3, the mapping $g = \inf\{h_{z_1} : i = 1, 2, \ldots, n\}$ is in $\bar{A}$, and satisfies the required condition.

Step 6. We finally show that $\bar{A} = \mathcal{C}(X,\mathbb{R})$. Let f be any point in $\mathcal{C}(X,\mathbb{R})$, and $\epsilon > 0$ be a real number. By step 5, for each x in X, there exists a mapping g_x in $\bar{A}$ such that $g_x(x) = f(x)$, and $g_x(y) \leq f(y) + \frac{\epsilon}{2}$ for all y in X. Since f and g_x are continuous at x, there exists an open neighbourhood U_x of x such that $g_x(y) \geq f(y) - \epsilon$ for all y in U_x. Since X is compact, X can be covered by a finite number of such neighbourhoods U_x, say, $U_{x_1}, U_{x_2}, \ldots, U_{x_m}$. Then the mapping $\phi = \sup\{g_{x_i} : i = 1, 2 \ldots, m\}$ is in $\bar{A}$ by step 3, and it is such that $f(y) - \epsilon \leq \phi(y) \leq f(y) + \epsilon$. Therefore,

$$d(f, \phi) = \sup\{|f(x) - \phi(x)| : x \in X\} \leq \epsilon$$

and this means that f is in the closure of $\bar{A}$, that is, f is in $\bar{A}$.

Theorem 8.3.3 is not true if we replace in it $\mathcal{C}(X,\mathbb{R})$ by $\mathcal{C}(X,\mathbb{C})$ as the following example indicates.

Example 8.3.4 Let D be the closed unit unite disk $\{z \in \mathbb{C} : |z| \leq 1\}$ in the complex plane $\mathbb{C}$. Then D is compact. Let A be the subset of $\mathcal{C}(D,\mathbb{C})$ containing those functions which are analytic in the open disk $\{z \in \mathbb{C} : |z| < 1\}$. It is easily seen that A is a subalgebra of $\mathcal{C}(D,\mathcal{C})$ that separates points in D, and contains the constant functions. But A is not dense in $\mathcal{C}(D,\mathbb{C})$. To prove this, we show that A is a closed set, and $A \neq \mathcal{C}(D,\mathbb{C})$. Observe that the function $g(z) = \bar{z}$ is continuous but not analytic, and so $g \in \mathcal{C}(D,\mathbb{C}) - A$. Now let $\langle f_n \rangle$ be sequence in A such that its pointwise limit is f. Then $\langle f_n \rangle$ converges to f uniformly on D, and consequently $\langle f_n \rangle$ converges to f uniformly on every compact subset of D. Then by Theorem 2 of Ahlfors [1] page 174, $f \in A$. Corollary 3.2.13 then shows that A is closed.

However, the following weaker result holds when $\mathcal{C}(X,\mathbb{R})$ in Stone-Weierstrass theorem is replaced by $\mathcal{C}(X,\mathcal{C})$.

Theorem 8.3.5 Let X be a compact metric space. If A is a subalgebra of $\mathcal{C}(X,\mathbb{C})$ that contains constant mappings, separates points of X, and for each mapping f in A its conjugate $\bar{f}$ is in A, then A is dense in $\mathcal{C}(X,\mathbb{C})$.

Proof. If $f \in A$, then the real valued mappings $\mathrm{Re}(f) = \frac{(f+\bar{f})}{2}$ and $\mathrm{Im}(f) = \frac{(f-\bar{f})}{2i}$ are also in A, and $f = \mathrm{Re}(f) + i\ \mathrm{Im}(f)$. Then the set A_0 of all real valued mappings in A with respect to the real scalars forms a subalgebra of $\mathcal{C}(X,\mathbb{R})$. We show that A_0 separates

points of X. Let x and y be distinct points of X. Since A separates points of X, there exists a mapping f in A such that $f(x) \neq f(y)$. This means that $\mathrm{Re}(f)(x) \neq \mathrm{Re}(f)(y)$, or $\mathrm{Im}(f)(x) \neq \mathrm{Im}(f)(y)$. Since $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are both in A_0, A_0 separates points of X. Next we show that A_0 contains a non-zero constant mapping. By hypothesis, A contains a non-zero constant mapping g. Since A is an algebra containing the conjugate of each of its element, $\bar{g} \in A$. Consequently, $g\bar{g} = |g|^2$ is a non-zero constant mapping in A_0. By Theorem 8.3.3, A_0 is dense in $\mathcal{C}(X, \mathbb{R})$. Since $A = A_0 + iA_0$, and $\mathcal{C}(X, \mathbb{C}) = \mathcal{C}(X, \mathbb{R}) + i\mathcal{C}(X, \mathbb{R})$, it follows that A is dense in $\mathcal{C}(X, \mathbb{C})$.

There are other classes of functions that approximate a continuous function. The corresponding theorems are corollaries of Stone-Weierstrass theorem. Our next theorem approximates a continuous real valued function by trigonometric polynomials. A trigonometric polynomial is a function p of the form

$$p(x) = \frac{\alpha_0}{2} + \sum_{r=1}^{n} (\alpha_r \cos(rx) + \beta_r \sin(rx)) \quad \text{for all } x \in \mathbb{R},$$

where $\alpha_0, \alpha_1, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}$. Note that p is a periodic function with period $2\pi : p(x + 2\pi) = p(x)$.

Theorem 8.3.6 (Weierstrass) Let f be a periodic real valued continuous functions defined on $\mathbb{R}$ with period 2π. Then for each real number $\epsilon > 0$, there is a trigonometric polynomial p such that $|f(x) - p(x)| < \epsilon$ for all x in $\mathbb{R}$.

Proof. Let $X = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ be the unit circle in the Euclidean plane $\mathbb{R}^2$. Then X is compact. There is a natural correspondence $g \to g^*$ between the set of periodic continuous functions defined on $\mathbb{R}$ with period 2π and $\mathcal{C}(X, \mathbb{R})$ given by $g^*(\cos x, \sin x) = g(x)$ for all x in $\mathbb{R}$. The set A of all functions p^* corresponding to trigonometric polynomial p is clearly a subalgebra of $\mathcal{C}(X, \mathbb{R})$ that separates points of X, and contains the constant functions. Therefore, by Stone-Weierstrass theorem, A is dense in $\mathcal{C}(X, \mathbb{R})$. Then there exists a trigonometric polynomial p such that $\|f^* - p^*\| < \epsilon$. Thus, for $x \in \mathbb{R}$

$$\begin{aligned} |f(x) - p(x)| &= |f^*(\cos x, \sin x) - p^*(\cos x, \sin x)| \\ &\leq \|f^* - p^*\| < \epsilon. \end{aligned}$$

Theorem 8.3.7 Any real valued continuous function on a compact subset E of $\mathbb{R}^n$ is the limit of a sequence of polynomials which converges uniformly in E.

Proof. Let A be the algebra of the restrictions to E of the polynomials in the n coordinates. Then A separates points of E since, for any two distinct points, at least one of the coordinates has distinct values. By Stone-Weierstrass Theorem, A is dense in $\mathcal{C}(X, \mathbb{R})$.

8.4 Separability of $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$

To establish the separability of the spaces $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$, we need in addition to Stone-Weirstrass theorem a little more background in normed linear spaces. The following Lemma plays a fundamental role in this development.

Lemma 8.4.1 Let X be a normed linear space, and $\{x_1, x_2, \ldots, x_n\}$ be a linearly independent set of vectors in X. Then there exists a real number $c > 0$ such that for every choice

of scalars $a_1, a_2, \ldots, a_n$,

$$\|a_1x_1 + a_2x_2 + \cdots + a_nx_n\| \geq c\,(|a_1| + \cdots + |a_n|)\,. \tag{8.6}$$

Proof. Let $a = |a_1| + \cdots + |a_n|$. If $a = 0$, then any value $c > 0$ will satisfy the conclusion of the theorem. So assume that $a > 0$. Then dividing the inequality 8.6 by a and denoting $b_i = \frac{a_i}{a}$, the inequality 8.6 is written equivalently as

$$\|b_1x_1 + \cdots + b_nx_n\| \geq c, \text{ where } \sum_{i=1}^{n} |b_i| = 1. \tag{8.7}$$

Therefore, it suffices to show that there exists a number $c > 0$ such that for every choice of scalars $b_1, b_2, \ldots, b_n$ such that $\sum_{i=1}^{n} |b_i| = 1$, the inequality (8.7) holds. Suppose on the contrary that there exists a sequence $\langle y_m \rangle$ of vectors such that

$$y_m = b_1^{(m)}x_1 + \cdots + b_n^{(m)}x_n \text{ where } \sum_{i=1}^{n} \left|b_i^{(m)}\right| = 1,$$

with $\|y_m\| \to 0$ as $m \to \infty$. Since $\sum_{i=1}^{m} |b_i^{(m)}| = 1$, $|b_i^{(m)}| \leq 1$. Therefore, the sequence of scalars in the j^{th} terms $\langle b_j^{(m)} \rangle_{m=1}^{\infty}$ is a bounded sequence, and consequently, by Bolzano-Weierstrass theorem, it has a convergent subsequence. Let b_1 be the limit of a convergent subsequence $\langle b_1^{(m_{k_1})} \rangle_{k_1=1}^{\infty}$ of $\langle b_1^{(m)} \rangle$. We select $\langle y_{m_{k_1}} \rangle_{k_1=1}^{\infty}$ as the corresponding sub sequence of $\langle y_m \rangle$. By the same argument, the sequence $\langle b_2^{(m_{k_1})} \rangle_{k_1=1}^{\infty}$ of the scalars in the second terms of the elements in $\langle y_{m_{k_1}} \rangle$ has a convergent subsequence converging to b_2, and let the corresponding subsequence of $\langle y_{m_{k_1}} \rangle_{k_1=1}^{\infty}$ be $\langle y_{(m_{k_1})_{k_2}} \rangle_{k_2=1}^{\infty}$; to simplify the notation, we write $y_{(m_{k_1})_{k_2}} = y_{m_{k_1,k_2}}$. The above process after n steps will give the subsequence $\langle y_{m_{k_1,k_2,\ldots,k_n}} \rangle_{k_n=1}^{\infty}$ of $\langle y_m \rangle$, where $y_{m_{k_1,k_2\ldots,k_n}} = \sum_{i=1}^{n} b_i^{(m_{k_1,\ldots,k_n})} x_i$, where $\sum_{i=1}^{n} |b_i^{(m_{k_1,\ldots,k_n})}| = 1$, and the sequence of scalars $\langle b_j^{(m_{k_1,\ldots,k_n})} \rangle_{k_n=1}^{\infty}$ converges to b_j as $k_n \to \infty$. Hence, the sequence $\langle y_{m_{k_1,k_2,\ldots,k_n}} \rangle \to y = \sum_{i=1}^{n} b_i x_i$ as $k_n \to \infty$, and it is easily seen that $\sum_{i=1}^{n} |b_i| = 1$. So that not all b_i can be zero. Since $\{x_1, \ldots, x_n\}$ is linearly independent, $y \neq 0$. On the other hand, the sequence $\langle y_{m_{k_1,\ldots,k_n}} \rangle \to y$ implies that $\|y_{m_{k_1,\ldots,k_n}}\| \to \|y\|$ by the continuity of the norm. Since $\|y_m\| \to 0$ by assumption, and $\langle y_{m_{k_1,k_2,\ldots,k_n}} \rangle$ is a subsequence of $\langle y_m \rangle$, $\|y_{m_{k_1,\ldots,k_2,k_n}}\| \to 0$. Therefore, $\|y\| = 0$ so that $y = 0$. This is a contradiction, and the proof is complete.

Theorem 8.4.2 Every finite dimensional subspace Y of a normed linear space X is complete.

Proof. Let $\langle y_m \rangle$ be a Cauchy sequence in Y. Let $\dim Y = n$, and $\{e_1, e_2, \ldots, e_n\}$ be any basis of Y. Then each y_m has a unique representation of the form $y_m = a_1^{(m)}e_1 + \cdots + a_n^{(m)}e_n$. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that $d(y_m, y_r) < \epsilon$ for all $m, r \geq p$. By Lemma 8.4.1, there exists a real number $\epsilon > 0$ such that

$$\epsilon > d(y_m, y_r) = \|y_m - y_r\| = \left\| \sum_{i=1}^{n} \left(a_i^{(m)} - a_i^{(r)}\right) e_i \right\| \geq c \sum_{i=1}^{n} \left|a_i^{(m)} - a_i^{(r)}\right| \quad \text{for all } m, r \geq p.$$

Therefore, $|a_i^{(m)} - a_i^{(r)}| < \frac{\epsilon}{c}$ for all $m, r \geq p$, and $i = 1, 2, \ldots, n$. This means that for each $i = 1, 2, \ldots, n$, the sequence of scalar $\langle a_i^{(m)} \rangle_{m=1}^{\infty}$ in the i^{th} terms is Cauchy sequence in $\mathbb{R}$ or

$\mathbb{C}$, and thus converges to a point $a_i \in \mathbb{R}$ (or $\mathbb{C}$). Let $y = a_1e_1 + \cdots + a_ne_n$. Clearly, $y \in Y$, and

$$d(y_m, y) = \left\| \sum_{i=1}^{n} \left(a_i^{(m)} - a_i \right) e_i \right\| \leq \sum_{i=1}^{n} \left| a_i^{(m)} - a_i \right| \|e_i\|.$$

Since $\langle a_i^{(m)} \rangle_{m=1}^{\infty}$ converges to a_i, $\langle y_m \rangle \to y$, and the proof is complete.

The following Corollary now follows from the above theorem.

Corollary 8.4.3 Every finite dimensional subspace Y of a normed linear space X is closed in X.

Note that an infinite dimensional subspace of a normed linear space need not be closed. For example, in the normed space $\mathcal{C}[0,1]$, let Y be the linear span of the set $\{x_0, x_1, \dots\}$ where $x_j(t) = t^j$ so that Y is the set of all polynomials. Y is not a closed set in $\mathcal{C}[0,1]$.

Definition 8.4.4 A subset A of a normed linear space X is said to be total if the linear span of A is dense in X. A sequence in X is said to be total in X if its set of points is total in X.

Theorem 8.4.5 Let X be a normed linear space. If there exists a total sequence in X, then X is separable. Conversely, if E is separable, then there exists a total sequence in X consisting of linearly independent vectors.

Proof. Let $\langle x_n \rangle$ be a total sequence, and let D be the set of all (finite) linear combinations $r_1x_1 + \cdots + r_nx_n$ with rational coefficients (when X is a complex vector space; by a "rational scalar" we mean a complex number $\alpha + i\beta$ with α and β rational). D is clearly a denumerable set. As by the definition, the set L of all linear combinations of the x_n's is dense in X, all we have to prove is that D is dense in L, and this follows from the inequality

$$\|(\lambda_1x_1 + \cdots + \lambda_nx_n) - (r_1x_1 + \cdots + r_nx_n)\| \leq \sum_{j=1}^{n} |\lambda_j - r_j| \, \|x_j\|,$$

and the fact that rationals are dense in $\mathbb{R}$ (respectively $\mathbb{C}$).

Conversely, suppose that X is separable. We can assume that X is infinite dimensional (otherwise any basis of X is already a finite total subset). Let $\langle x_n \rangle$ be an infinite dense sequence of vectors of X. We define, by induction, a subsequence $\langle x_{n_k} \rangle$ having the property that it consists of linearly independent vectors, and that for an $m \leq n_k, x_m$ is a linear combination of $x_{n_1}, x_{n_2}, \dots, x_{n_k}$. To do this, we merely take n_1 the first index for which $x_k \neq 0$, and for n_{k+1} the smallest index $m > n_k$ such that x_m is not in the subspace V_k generated by $x_{n_1}, \dots, x_{n_k}$; such an index exists, otherwise as V_k is closed by corollary 8.4.3, V_k would contain the closure X of the set of all the $x_n's$ contrary to assumption. It is then clear that $\langle x_{n_k} \rangle$ has the required properties, and is obviously, by construction, a total sequence.

Now we are in a position to prove the main theorem of this section.

Theorem 8.4.6 Let (X, d) be a compact metric space. Then the metric spaces $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(X, \mathbb{C})$ are separable.

Proof. First we show that $\mathcal{C}(X, \mathbb{R})$ is separable. Since X is compact, it satisfies the second axiom of countablity. Therefore, there exists a denumerable open base $\{U_n\}$ for X. For all n, let $g_n : X \to \mathbb{R}$ be the mapping defined by $g_n(x) = d(x, X - U_n)$. Then g_n is continuous. For each n, $A_n = \{g_n^{\alpha} : \alpha \text{ is a non-negative integer}\}$ is a denumerable set, and also the set

H_n of monomials $g_1^{\alpha_1} g_2^{\alpha_2} \dots g_n^{\alpha_n}$ is denumerable. Let A be the vector space generated by the denumerable set $H = \bigcup_{n=1}^{\infty} H_n$. Then A is the subalgebra of $\mathcal{C}(X, \mathbb{R})$ generated by H. We show that A is dense in $\mathcal{C}(X, \mathbb{R})$. It is obvious that A contains constant mappings. Also, if $x \neq y$, then there is a U_n such that $x \in U_n$ and $y \notin U_n$. Therefore, by the definition, $g_n(x) \neq 0$ and $g_n(y) = 0$. Thus, A separates the points of X. By the Stone-Weiertrass theorem, A is dense in $\mathcal{C}(X, \mathbb{R})$. Thus, H is total, and $\mathcal{C}(X, \mathbb{R})$ is separable by Theorem 8.4.5.

To show that $\mathcal{C}(X, \mathbb{C})$ is separable let $\{g_n : n = 1, 2, \dots\}$ be a countable and dense subset of $\mathcal{C}(X, \mathbb{R})$. Clearly, the subset $\{g_n + ig_m : n, m = 1, 2, \dots\}$ is countable. We show that it is dense in $\mathcal{C}(X, \mathbb{C})$. Let $f \in \mathcal{C}(X, \mathbb{C})$, and $\epsilon > 0$ be a real number. Then $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are in $\mathcal{C}(X, \mathbb{R})$. So there exist positive integer n and m such that

$$\|\mathrm{Re}(f) - g_n\| < \frac{\epsilon}{2} \quad \text{and} \quad \|\mathrm{Im}(f) - g_m\| < \frac{\epsilon}{2}.$$

Thus,

$$\|f - (g_n + ig_m)\| = \|\mathrm{Re}(f) - g_n + i(\mathrm{Im}(f) - g_m)\| < \epsilon.$$

Therefore, the set $\{g_n + ig_m : n, m = 1, 2 \dots\}$ is dense in $\mathcal{C}(X, \mathbb{C})$.

The metric space $\mathcal{C}[a, b]$ is a particular case of the above theorem, and thus it is separable.

8.5 Existence of Functions Continuous Everywhere and Differentiable Nowhere

Consider the metric space $\mathcal{C}[0, 1]$, where sup metric is denoted by d. Let $A = \{f \in \mathcal{C}[0, 1] : f$ has right hand derivative at some $x \in [0, 1]\}$. Let n be a positive integer, and

$$E_n = \left\{ f \in \mathcal{C}[0,1] : \frac{|f(x+h) - f(x)|}{h} \leq n \quad \text{for all } h \in \left]0, \tfrac{1}{n}\right[, \text{ for some } x \in [0, 1 - \tfrac{1}{n}] \right\}.$$

It is easily seen that $A \subseteq \bigcup_{n=1}^{\infty} E_n$. We show that $\bigcup_{n=1}^{\infty} E_n$ is of first category. For this we show that E_n is a nowhere dense set, or that

$$\overline{((\bar{E}_n)')} = X. \tag{8.8}$$

We show that $\bar{E_n} = E_n$, that is, E_n is a closed set. Suppose that $g \in \bar{E_n}$. Then there exists a sequence $\langle f_j \rangle$ in E_n which converges to g with respect to the metric d. Since each f_j is in E_n,

$$\frac{|f_j(x_j + h) - f_j(x_j)|}{h} \leq n \quad \text{for all } h \in \left]0, \tfrac{1}{n}\right[, \text{ for some } x_j \in [0, 1 - \tfrac{1}{n}].$$

Since $\langle x_j \rangle$ is a sequence in $[0, 1 - \frac{1}{n}]$, it is bounded, and by Bolzano-Weierstrass theorem, it has a subsequence $\langle x_{j_k} \rangle$ which converges to a point $x_0 \in [0, 1 - \frac{1}{n}]$. Since any subsequence of a convergent sequence converges to the same limit, the subsequence $\langle f_{j_k} \rangle$ also converges to g, and this convergence is uniform on $[0, 1]$. Also since g is continuous, $\lim_{k \to \infty} g(x_{j_k}) = g(\lim_{k \to \infty} x_{j_k}) = g(x_0)$. By the triangle inequality

$$\begin{aligned} |g(x_0 + h) - g(x_0)| \leq{}& |g(x_0 + h) - f_{j_k}(x_0 + h)| + |f_{j_k}(x_0 + h) - f_{j_k}(x_0)| \\ &+ |f_{j_k}(x_0) - f_{j_k}(x_{j_k})| + |f_{j_k}(x_{j_k}) - g(x_{j_k})| + |g(x_{j_k}) - g(x_0)|, \end{aligned}$$

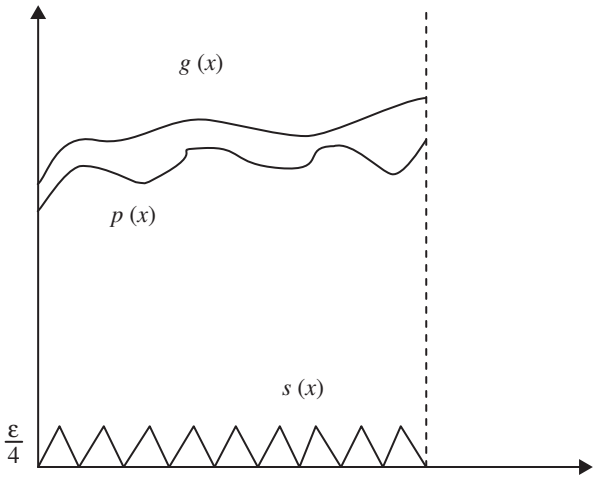


Fig. 8.1

it follows that $\frac{|g(x_0+h)-g(x_0)|}{h} \leq n$. Since $x_0 \in [0, 1-\frac{1}{n}]$, $g \in E_n$. Therefore, to prove (8.8), we need to show that $\overline{(E_n')} = X$, that is, E_n' is dense in X. Now

$$E_n' = \left\{ f \in \mathcal{C}[0,1] : \frac{|f(x+h)-f(x)|}{h} > n \quad \text{for some } h \in]0, \tfrac{1}{n}[\text{ and all } x \in [0, 1-\tfrac{1}{n}] \right\}.$$

Let $g \in \mathcal{C}[0,1]$, and $\epsilon > 0$ be a real number. By Weierstrass approximation theorem, there exists a polynomial p such that $d(g,p) < \frac{\epsilon}{4}$. Now let $s(x)$ be sawtooth function defined on [0,1] (Fig. 8.1) whose slope is greater than $n + \max\{|p'(x)| : x \in [0,1]\}$ and $|s(x)| < \frac{\epsilon}{4}$. Then $d(g,(p+s)) < \epsilon$, and $p+s \in E_n'$. Thus, E_n' is dense in X. Hence $\bigcup_{n=1}^{\infty} E_n$ is of first category, and so A is of first category. Since $\mathcal{C}[0,1] = A \cup A'$, and $\mathcal{C}[0,1]$ is of second category, then A' is of second category. Therefore, A' is non-empty. Since A' is the set of all functions in $\mathcal{C}[0,1]$ that does not possess right hand derivative, existence of a function continuous everywhere and differentiable nowhere is proved.

EXERCISE

1. Prove that the polynomials are dense in $\mathcal{C}(X, \mathbb{R})$ for any closed and bounded subspace X of the real line $\mathbb{R}$.

2. Let A be any set, and $\mathbb{B}(A, \mathbb{R})$ be the set of all bounded mappings with sup metric. Show that the mapping $g : \mathbb{B}(A, \mathbb{R}) \to \mathbb{R}$ defined by

$$g(f) = \sup\{f(x) : x \in A\}$$

is continuous.

3. Let (X, d) be a metric space with bounded metric d, and $\mathcal{C}_b(X, \mathbb{R})$ be the Banach space with sup metric. For each $x \in X$, let $f_x : X \to \mathbb{R}$ be the bounded and continuous mapping defined by $f_x(y) = d(x, y)$. Show that the mapping $g : X \to \mathcal{C}_b(X, \mathbb{R})$ defined by $g(x) = f_x$ is an isometry of X onto a subspace of Banach space $\mathcal{C}_b(X, \mathbb{R})$.

4. Show that the Banach space $\mathcal{C}_b(\mathbb{R}, \mathbb{R})$ of continuous and bounded mapping with sup norm is not separable.

5. Let $I = [0, 1]$ be the subspace of $\mathbb{R}$. Let $x_1, \ldots, x_n$ be n distinct points in I. For each $k = 1, 2, \ldots, n$, let $f_k : I \to \mathbb{R}$ be the function defined by $f_k(x) = |x - x_k|$. Show that $f_1, f_2, \ldots, f_n$ are linearly independent in $\mathcal{C}(I, \mathbb{R})$.

Chapter 9

Fixed Point Theorems and Their Applications

Let X be a set, and $f : X \to X$ be a mapping. A point x in X is said to be a fixed point of f if $f(x) = x$. There are a number of problems in mathematics whose solutions are fixed points of certain mappings. Sometimes it is not possible to find a fixed point of a mapping exactly. Approximation methods are applied to get a value which is near to the required fixed point. Newton's method in numerical analysis can be considered such an approximation method. Some of these problems will be discussed in this chapter.

9.1 Fixed Point Theorems

Let X be a set, and $f : X \to X$ be a mapping. Then f need not have a fixed point. However, if we impose a certain mathematical structure on X and f, then f may have a fixed point. Our first attempt in this direction is the following.

Theorem 9.1.1 Every continuous function $f : [a, b] \to [a, b]$ has atleast one fixed point.

Proof. If a or b is a fixed point of f, then there is nothing to prove. So we assume that $f(a) \neq a$ and $f(b) \neq b$.

Let us define a function $g : [a, b] \to [a, b]$ by setting $g(x) = f(x) - x$. Then g is obviously a continuous function. Since $f(a)$ is in $[a, b]$, $f(a) > a$. Similarly, $f(b) < b$. This means that $g(a) > 0$, and $g(b) < 0$. Since g is continuous on $[a, b]$, by Corollary 7.5.4, there is a point x_0 in $]a, b[$ such that $g(x_0) = 0$, that is, $f(x_0) = x_0$. Thus, x_0 is a fixed point of f.

The above theorem graphically may be explained as in Fig. 9.1. A fixed point of f is a point on the line $y = x$ where the graph of f intersects. If we assume that a and b are not fixed points of f, then $f(a) > a$, and $f(b) < b$. That means that the point $(a, f(a))$ is above the line $y = x$,

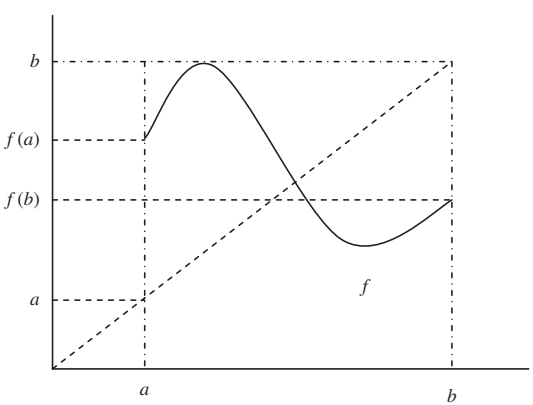


Fig. 9.1

and the point $(b, f(b))$ is below the line $y = x$. If we draw the graph of continuous function f from $(a, f(a))$ to $(b, f(b))$, then we will cross the line $y = x$ at some point, and this point is a fixed point of f.

Definition 9.1.2 Let (X, d) be a metric space. A mapping $f : X \to X$ is said to be a contraction mapping if there exists a real number λ, $0 \leq \lambda < 1$, such that

$$d(f(x), f(y)) \leq \lambda d(x, y) \quad \text{for all } x, y \text{ in } X.$$

It is stressed that λ in the above definition is independent of x and y in X.

The following result follows directly from the definition of a contraction mapping.

Theorem 9.1.3 Every contraction mapping is uniformly continuous.

Theorem 9.1.4 (Banach fixed point) Every contraction mapping on a complete metric space has a unique fixed point.

Proof. Let (X, d) be a complete metric space, and f be a contraction on X. Then there exists a real number λ, $0 \leq \lambda < 1$ such that $d(f(x), f(y)) \leq \lambda d(x, y)$ for all x, y in X. Let x_0 be any arbitrary point of X. We define a sequence $\langle x_n \rangle$ inductively as follows:

$$\begin{aligned} x_1 &= f(x_0), \\ x_{n+1} &= f(x_n) \text{ for } n \geq 1. \end{aligned}$$

We show that $\langle x_n \rangle$ is a Cauchy sequence in X. Let $\epsilon > 0$ be a real number, and $n > m$. Then

$$\begin{aligned} d(x_n, x_m) &= d(f(x_{n-1}), f(x_{m-1})) \\ &\leq \lambda d(x_{n-1}, x_{m-1}) \\ &\leq \lambda^2 d(x_{n-2}, x_{m-2}) \\ &\vdots \\ &\leq \lambda^m d(x_{n-m}, x_0). \end{aligned}$$

Now

$$\begin{aligned} d(x_{n-m}, x_0) &\leq d(x_{n-m}, x_{n-m-1}) + d(x_{n-m-1}, x_{n-m-2}) + \cdots + d(x_1, x_0) \\ &\leq \lambda^{n-m-1} d(x_1, x_0) + \lambda^{n-m-2} d(x_1, x_0) + \cdots + d(x_1, x_0) \\ &< \frac{d(x_1, x_0)}{1 - \lambda}. \end{aligned}$$

Therefore, $d(x_n, x_m) < \frac{d(x_1, x_0)\lambda^m}{(1-\lambda)}$. Since $0 \leq \lambda < 1$, there exists a $p \in \mathbb{Z}_+$ such that $\frac{d(x_1, x_0)\lambda^m}{(1-\lambda)} < \epsilon$ for all $m \geq p$. This proves that $\langle x_n \rangle$ is a Cauchy sequence in X. Since X is complete, $\langle x_n \rangle$ converges to a point x in X. Now

$$\begin{aligned} d(f(x), x) &\leq d(f(x), x_n) + d(x_n, x) \\ &\leq \lambda d(x, x_{n-1}) + d(x_n, x). \end{aligned}$$

Since $\langle x_n \rangle$ converges to x, there exists a $k \in \mathbb{Z}_+$ such that $\lambda d(x, x_{n-1}) + d(x_n, x) < \epsilon$ for all $n \geq k$. Thus, $d(f(x), x) < \epsilon$. Since ϵ is arbitrary real number, $d(f(x), x) = 0$, that is,

$f(x) = x$, and x is a fixed point of f. To show that x is a unique fixed point, let y be another fixed point of f. Then

$$d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y)$$

which gives $d(x, y) = 0$, so that $x = y$, a contradiction. This completes the proof.

The reader should observe that the proof of the above theorem is constructive; the existence of fixed point is established by constructing a sequence that converges to that fixed point. This theorem is an example of a very wide class of similar but more general results concerning the existence of fixed points of mappings. We emphasise that such theorems play very important part in modern mathematics. Some applications of the above theorem will be covered in this chapter.

The following trivial, yet somewhat surprising extension, of the above theorem is frequently used. In this theorem, the mapping f is not assumed to be even continuous. For a mapping f, we write $f^n = f \circ f \circ \cdots \circ f$, the composition of f n-times.

Theorem 9.1.5 Let (X, d) be a complete metric space, and $f : X \to X$ be a mapping such that f^n is a contraction mapping for some $n \in \mathbb{Z}_+$, that is, for some λ, $0 \leq \lambda < 1$

$$d(f^n(x), f^n(y)) \leq \lambda d(x, y) \text{ for all } x, y \text{ in } X.$$

Then f has a unique fixed point.

Proof. By Theorem 9.1.4, the contraction f^n has a unique fixed point x in X. Then

$$d(f(x), x) = d(f^{n+1}(x), f^n(x)) \leq \lambda d(f(x), x)$$

So that $f(x) = x$, and x is a fixed point of f, and it is easily seen to be unique.

Example 9.1.6 The condition of completeness in Theorem 9.1.4 can not be dropped. The mapping $f :]0, \frac{1}{3}] \to]0, \frac{1}{3}]$ defined by $f(x) = x^2$ is a contraction, but it has no fixed point in $]0, \frac{1}{3}]$.

Theorem 9.1.7 Let $f : [a, b] \to [a, b]$ be differentiable. Then f is a contraction on $[a, b]$ if and only if there exists a real number $k < 1$ with $|f'(x)| \leq k$ for all x in the open interval $]a, b[$.

Proof. Let f be a contraction on $[a, b]$. Then there exists a real number k, $0 \leq k < 1$ such that $|f(x) - f(y)| \leq k|x - y|$ for all x, y in $[a, b]$. In particular, $|f(x + \Delta x) - f(x)| \leq k|\Delta x|$ for all x and $x + \Delta x$ in $[a, b]$. Therefore,

$$\left| \frac{f(x + \Delta x) - f(x)}{\Delta x} \right| \leq k.$$

This means that $\lim_{\Delta x \to 0} |\frac{f(x+\Delta x)-f(x)}{\Delta x}| \leq k$, so that $|f'(x)| \leq k$. Conversely, assume that $|f'(x)| \leq k$, where the real number $k < 1$ for all x in $]a, b[$. For $y \neq x$, the Lagrange's theorem mean value of calculus gives that

$$\frac{f(x) - f(y)}{x - y} = f'(c).$$

for some c between x and y. Therefore,

$$\left| \frac{f(x) - f(y)}{x - y} \right| = |f'(c)| \leq k$$

that is, $|f(x) - f(y)| \leq k|x - y|$, and f is a contraction.

Theorem 9.1.8 Let (X, d) be a compact metric space, and $f : X \to X$ be a contractive mapping, that is, $d(f(x), f(y)) < d(x, y)$ for all distinct x, y in X. Then f has a unique fixed point in X.

Proof. We define $g : X \to \mathbb{R}$ by $g(x) = d(x, f(x))$ for all x in X. Then the continuity of g follows by the inequality:

$$\begin{aligned} |g(x) - g(y)| &= |d(x, f(x)) - d(y, f(y))| \\ &\leq |d(x, f(x)) - d(f(x), y)| + |d(f(x), y) - d(y, f(y))| \\ &\leq d(x, y) + d(f(x), f(y)) \\ &< 2d(x, y). \end{aligned}$$

Since X is compact, g attains it infimum at some point x in X. We show that x is a fixed point of f. Suppose that $y = f(x) \neq x$. Then $g(x) \leq g(y) \leq g(f(x))$. Therefore, by the definition of g

$$d(x, f(x)) \leq d(f(x), f(f(x))) < d(x, f(x)).$$

a contradiction. Thus, x is a fixed point of f. The uniqueness is obvious.

Example 9.1.9 Let $X = [1, \infty[$, and $T : X \to X$ be a mapping defined by $T(x) = x + \frac{1}{x}$. Then $|T(x) - T(y)| < |x - y|\frac{(xy-1)}{xy}$ for all distinct x and y in X, and so T is contractive. Since $\lim_{y\to x} \frac{(xy-1)}{xy} = 1$, when x is large, T is not a contraction mapping.

9.2 Applications

9.2.1 Estimate of Error in the Calculation of a Fixed Point

Sometimes in an application, we do not calculate the exact fixed point of a contraction mapping, and some point which is close to the actual fixed point is enough. In that case, it is possible to estimate the error (the distance between the point selected and the actual fixed point). More specifically, in the proof of Banach's fixed point theorem 9.1.4, we constructed a sequence $\langle x_n \rangle$ which satisfied the inequality

$$d(x_n, x_m) \leq \frac{d(x_1, x_0)\lambda^m}{(1 - \lambda)}, \quad (0 \leq \lambda < 1).$$

Keeping m fixed and taking limit as n tends to infinity in the above inequality, we get

$$d(x_m, x) \leq \frac{d(x_1, x_0)\lambda^m}{(1 - \lambda)}$$

where x is the fixed point of the contraction mapping. This shows that the error in taking the point x_m instead of x does not exceed $\frac{d(x_1,x_0)\lambda^m}{(1-\lambda)}$. If this error is to be less than a preassined real number $\epsilon > 0$, then

$$m > \frac{\log\left(\frac{d(x_1,x_0)}{\epsilon(1-\lambda)}\right)}{\log(\frac{1}{\lambda})}.$$

This gives the number of iterations m needed to bring the point x_m within ϵ distance of the actual fixed point.

Theorem 9.1.4 is key to prove the existence of a unique solution of first order differential equation.

Theorem 9.2.2 (Picard) If $f(x, y)$ and $\frac{\partial f}{\partial y}$ are continuous in a closed rectangle $D = \{(x, y) \in \mathbb{R}^2 : a_1 \leq x \leq a_2, \text{ and } b_1 \leq y \leq b_2\}$, and if (x_0, y_0) is an interior point of D, then the differential equation

$$\frac{dy}{dx} = f(x, y) \tag{9.1}$$

has a unique solution $y = g(x)$ which passes through the point (x_0, y_0).

Proof. Since $f(x, y)$ and $\frac{\partial f}{\partial y}$ are continuous on D, and D is compact, $f(x, y)$ and $\frac{\partial f}{\partial y}$ are bounded, and so there exist real numbers k and s such that

$$|f(x, y)| \leq k \quad \text{and} \quad \left|\frac{\partial f(x, y)}{\partial y}\right| \leq s$$

for all (x, y) in D. Now if (x, y_1) and (x, y_2) be any points in D, then, by the Lagrange's mean value theorem of calculus

$$|f(x, y_1) - f(x, y_2)| = |y_1 - y_2| \left|\frac{\partial f(x, y_1) + \theta(y_1 - y_2)}{\partial y}\right|$$

for some θ such that $0 < \theta < 1$. Thus,

$$|f(x, y_1) - f(x, y_2)| \leq s|y_1 - y_2| \quad \text{for all } (x, y_1) \text{ and } (x, y_2) \text{ in } D.$$

Now if $y = g(x)$ is a solution of (9.1) such that $g(x_0) = y_0$, then integrating (9.1) from x_0 to x yields

$$g(x) - g(x_0) = \int_{x_0}^{x} f(t, g(t))\, dt$$

or

$$g(x) = y_0 + \int_{x_0}^{x} f(t, g(t))\, dt. \tag{9.2}$$

Conversely, if $y = g(x)$ satisfies (9.2), then it follows that $g(x_0) = y_0$. Thus, $y = g(x)$ such that $y_0 = g(x_0)$ is a solution of (9.1) if and only if it satisfies the integral equation (9.2). Therefore, it is sufficient to prove that (9.2) has a unique solution. Now we can choose a real number $\delta > 0$ such that $s\delta < 1$ and the closed rectangle $D' = \{(x, y) \in \mathbb{R}^2 : |x - x_0| \leq \delta$ and $|y - y_0| < k\delta\}$ is contained in D. Now, let B be the set of all real valued mappings $y = g(x)$ which are defined and continuous in the closed interval $|x - x_0| \leq \delta$, and such that $|g(x) - y_0| \leq k\delta$. We show that B is a closed subspace of the complete metric space of continuous functions $\mathcal{C}[x_0 - \delta, x_0 + \delta]$ with sup metric d. Let ψ in $\mathcal{C}[x_0 - \delta, x_0 + \delta]$ be a limit point of B. Then there exists a sequence $\langle g_n \rangle$ in B which converges in $\mathcal{C}[x_0 - \delta, x_0 + \delta]$ to ψ. Let $\epsilon > 0$ be a real number. Then there exists a $p \in \mathbb{Z}_+$ such that $d(g_n, \psi) < \epsilon$ for all $n \geq p$. This means that $|g_n(x) - \psi(x)| < \epsilon$ for all $n \geq p$ and $x \in [x_0 - \delta, x_0 + \delta]$. Now, by the triangle inequality,

$$\begin{aligned} |\psi(x) - y_0| &\leq |\psi(x) - g_p(x)| + |g_p(x) - y_0| \\ &< \epsilon + k\delta. \end{aligned}$$

Since ϵ is arbitrary, $|\psi(x) - y_0| \leq k\delta$ for all $x \in [x_0 - \delta, x_0 + \delta]$. Thus, ψ is in B, and B is a closed set. Therefore, B is complete. In the next step, we define a mapping $T : B \to B$ by

$$T(\psi) = \phi, \text{ where } \phi(x) = y_0 + \int_{x_0}^{x} f(t, \psi(t))\, dt.$$

It is easily seen that ϕ is a continuous function on the closed interval $[x_0 - \delta, x_0 + \delta]$, and

$$|\phi(x) - y_0| = \left| \int_{x_0}^{x} f(t, \psi(t)) dt \right| \leq k\delta.$$

Thus, $\phi \in B$, and T is indeed a mapping from B into itself. Now

$$\begin{aligned} |\phi_1(x) - \phi_2(x)| &= \left| \int_{x_0}^{x} [f(t, \psi_1(t)) - f(t, \psi_2(t))]\ dt \right| \\ &\leq s\delta \sup\{|\psi_1(t) - \psi_2(t)\}. \end{aligned}$$

Since $s\delta < 1$, the above inequality shows that T is a contraction. Therefore, by Banach fixed point theorem (Theorem 9.1.4), the contraction mapping T has a unique fixed point g in B, that is, $T(g) = g$. This means that there is a unique function g in B such that

$$g(x) = y_0 + \int_{x_0}^{x} f(t, g(t))\, dt,$$

and the proof is complete.

Example 9.2.3 The condition of continuity of $\frac{\partial f}{\partial y}$ in the above theorem can not be dropped. For example, consider the differential equation

$$\frac{dy}{dx} = y^{\frac{2}{3}}, \quad y(0) = 0.$$

Then the function which is identically zero: $g(x) = 0$ for all x, satisfies the differential equation, and the initial condition $g(0) = 0$. The another solution obtained by integrating is $y = \frac{x^3}{27}$. Note that for the function $f(x, y) = y^{\frac{2}{3}}$, $\frac{\partial f}{\partial y}$ does not exist at $(0, 0)$.

An integral equation of the form

$$f(x) = g(x) + \lambda \int_a^b k(x, y) f(y)\, dy \quad \text{for all } x \text{ in } [a, b] \tag{9.3}$$

is known as a Fredholm integral equation. Here $f(x)$ and $g(x)$ are real valued functions defined on $[a, b]$, $k(x, y)$, known as kernel, is a real valued mapping defined on the square $[a, b] \times [a, b]$ in $\mathbb{R}^2$, and λ is a real number called parameter. The meaning of solving the integral equation (9.3) is to find a function $f(x)$ which satisfies the equation (9.3). In general, Fredholm integral equation can be considered on various metric spaces of functions. For our present purpose, we shall consider the metric space $\mathcal{C}[a, b]$ of continuous real valued functions defined on $[a, b]$ with sup metric.

Theorem 9.2.4 Let g be a function in $\mathcal{C}[a, b]$, and k be a continuous real valued mapping defined on the square $[a, b] \times [a, b]$ with subspace metric of $\mathbb{R}^2$. Then there exists a real number λ such that the Fredholm equation

$$f(x) = g(x) + \lambda \int_a^b k(x, y) f(y)\, dy \tag{9.4}$$

has a unique solution in $\mathcal{C}[a, b]$.

Proof. Since $[a,b]\times[a,b]$ is compact, and k is continuous on $[a,b]\times[a,b]$, k is bounded. Thus, there exists a real number M such that

$$|k(x,y)|\le M \text{ for all } (x,y)\in[a,b]\times[a,b].$$

We define a mapping $T:\mathcal{C}[a,b]\to\mathcal{C}[a,b]$ by setting

$T(\phi)=\psi$, where $\psi(x)=g(x)+\lambda\int_a^b k(x,y)\phi(y)dy$, and λ is a real number.
It is easily seen that ψ is continuous on $[a,b]$. Thus, T is indeed a mapping from $\mathcal{C}[a,b]$ into itself. Now

$$\begin{aligned}
d(\psi_1,\psi_2) &= \sup_{x\in[a,b]}\{|\psi_1(x)-\psi_2(x)|\}\\
&= |\lambda|\sup_{x\in[a,b]}\left\{\left|\int_a^b k(x,y)(\phi_1(y)-\phi_2(y))\,dy\right|\right\}\\
&\le |\lambda|M\sup_{x\in[a,b]}\left[\int_a^b|\phi_1(y)-\phi_2(y)|\,dy\right]\\
&\le |\lambda|M\sup_{y\in[a,b]}\{|\phi_1(y)-\phi_2(y)|\}(b-a)\\
&= |\lambda|M(b-a)d(\phi_1,\phi_2).
\end{aligned}$$

Thus, T is a contraction mapping if and only if $|\lambda|<\frac{1}{M(b-a)}$. Choose any λ satisfying $|\lambda|<\frac{1}{M(b-a)}$. Since $\mathcal{C}[a,b]$ is complete, by Theorem 9.1.4, the mapping T has a unique fixed point f in $\mathcal{C}[a,b]$, that is, $T(f)=f$. So there is a unique f in $\mathcal{C}[a,b]$ satisfying

$$f(x)=g(x)+\lambda\int_a^b k(x,y)f(y)\,dy.$$

This completes the proof.

Indeed, if we follow the procedure of construction of fixed point in Theorem 9.1.4, then starting from any function f_0 in $\mathcal{C}[a,b]$ one can construct a sequence $\langle f_n\rangle$ in $\mathcal{C}[a,b]$ inductively

$$f_1=T(f_0),\quad \text{and}\quad f_{n+1}=T(f_n),\quad n\ge 1,$$

that is, $f_{n+1}(x)=g(x)+\lambda\int_a^b k(x,y)f_n(y)dy$. This sequence $\langle f_n\rangle$ will converge to the required fixed point f in $\mathcal{C}[a,b]$.

An integral equation of the form

$$f(x)=g(x)+\lambda\int_a^x k(x,y)f(y)dy\quad\text{for all } x \text{ in } [a,b]$$

is called a Volterra integral equation. The next theorem shows that there is a unique solution of this integral equation in $\mathcal{C}[a,b]$.

Theorem 9.2.5 Let g be a function in $\mathcal{C}[a,b]$, and k be a continuous real valued mapping defined in the triangular region

$$D=\{(x,y)\in\mathbb{R}^2 : a\le x\le b \text{ and } a\le y\le x\}$$

in $\mathbb{R}^2$.

Then the Volterra integral equation

$$f(x) = g(x) + \lambda \int_a^x k(x,y) f(y)\, dy \quad \text{for all } x \text{ in } [a,b] \tag{9.5}$$

has a unique solution in $\mathcal{C}[a,b]$.

Proof. k is bounded, and so there exists a real number M such that $|k(x,y)| \leq M$ for all (x,y) in D. We define a mapping $T : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ by setting

$$T(\phi) = \psi, \text{ where } \psi(x) = g(x) + \lambda \int_a^x k(x,y)\phi(y)\, dy.$$

It is easily seen that T is indeed a mapping from $\mathcal{C}[a,b]$ into itself. Now

$$\begin{aligned} |T(\phi_1)(x) - T(\phi_2)(x)| &= \left| \lambda \int_a^x k(x,y)(\phi_1(y) - \phi_2(y))\, dy \right| \\ &\leq |\lambda| M d(\phi_1, \phi_2)(x-a) \quad \text{for all } x \text{ in} [a,b]. \end{aligned} \tag{9.6}$$

We show by induction the following inequality

$$|T^m(\phi_1)(x) - T^m(\phi_2)(x)| \leq \frac{|\lambda^m| M^m (x-a)^m d(\phi_1, \phi_2)}{m!} \text{ for all } x \text{ in } [a,b].$$

For $m = 1$, it becomes the statement (9.6). Now

$$\begin{aligned} T^{m+1}(\phi_1)(x) - T^{m+1}(\phi_2)(x)| &\leq |\lambda| M \int_a^x |T^m(\phi_1)(y) - T^m(\phi_2)(y)|\, dy \\ &\leq |\lambda| M \int_a^x \frac{|\lambda|^m M^m (y-a)^m d(\phi_1, \phi_2)}{m!} dy \\ &= \frac{|\lambda|^{m+1} M^{m+1} (x-a)^{m+1} d(\phi_1, \phi_2)}{(m+1)!}. \end{aligned}$$

Therefore,

$$|T^m(\phi_1)(x) - T^m(\phi_2)(x)| \leq \frac{|\lambda|^m M^m (b-a)^m d(\phi_1, \phi_2)}{m!} \quad \text{for all } x \text{ in } [a,b].$$

So that $d(T^m(\phi_1), T^m(\phi_2)) \leq \alpha d(\phi_1, \phi_2)$ where $\alpha = \frac{|\lambda|^m M^m (b-a)^m}{m!}$. For sufficiently large values of m, $\alpha < 1$, and T^m is a contraction mapping. By Theorem 9.1.5, the contraction mapping T has a unique fixed point f in $\mathcal{C}[a,b]$, and this f is the unique solution of Volterra integral equation (9.5).

Next we prove the implicit function theorem. First we define the meaning of an implicit function. Let

$$F(x_1, x_2, \ldots, x_n, y) = 0 \tag{9.7}$$

be a relation among $n+1$ variables $x_1, x_2, \ldots, x_n, y$, and let

$$x_1 = a_1,\ x_2 = a_2,\ \ldots,\ x_n = a_n$$

be a set of values such that the equation

$$F(a_1, a_2, \ldots, a_n, y) = 0 \tag{9.8}$$

is satisfied by atleast one value of y , that is, equation (9.8) in y has has at least one root. We may consider a function $y = \phi(x_1, x_2, \ldots, x_n)$ defined in a domain D such that at any point $(x_1, x_2, \ldots, x_n)$ in D, $y = \phi(x_1, x_2, \ldots, x_n)$ is a root of equation (9.7), that is, $F(x_1, x_2, \ldots, x_n, \phi(x_1, x_2, \ldots, x_n)) = 0$. We say that $y = \phi(x_1, x_2, \ldots, x_n)$ is an implicit function with domain D defined by the relation (9.7).

For example, the relation

$$F(x, y) = x^2 + y^2 - 4 = 0$$

defines infinite number of functions of the form $y = \phi(x)$ in the domain $D = \{x : x^2 \leq 4\}$. Two of them are $y = \sqrt{4 - x^2}$ and $y = -\sqrt{4 - x^2}$; The others can be formed as follows: let J be any non-empty proper subset of D. Then

$$y = \begin{cases} \sqrt{4 - x^2} & \text{if } x \in J, \\ -\sqrt{4 - x^2} & \text{if } x \in D - J. \end{cases}$$

As a further illustration, consider the relation

$$F(x, y) = y^5 - y - x = 0.$$

From the above relation, although x can be expressed explicity in terms of y by $x = y^5 - y$, no algebraic expression of y in terms of x can be found. Implicit function theorem tells us under what conditions a given relation like (9.7) defines y as a function of x in a domain. In order to keep the presentation simple, we consider the relations of the form $F(x, y) = 0$ to derive our next result, which is popularly known as the implicit function theorem.

Theorem 9.2.6 (Implicit Function) Let D be an open set in $\mathbb{R}^2$ containing a point (x_0, y_0), and $F : D \to \mathbb{R}$ be a mapping such that

(i) $F(x, y)$ is continuous in D,

(ii) $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ are both continuous in D, and

(iii) $F(x_0, y_0) = 0$, and $\frac{\partial F}{\partial y}|_{(x_0, y_0)} \neq 0$.

Then there exists a rectangle $R = \{(x, y) : |x - x_0| < h,\ |y - y_0| < k\}$ contained in D, and a function $g : [x_0 - h, x_0 + h] \to [y_0 - k, y_0 + k]$ such that
(i) $f(x, g(x)) = 0$ for all x in $[x_0 - h, x_0 + h]$, and
(ii) g is differentiable in $[x_0 - h, x_0 + h]$, and its derivative

$$g'(x) = -\left(\frac{\partial F}{\partial x}\right) \Big/ \left(\frac{\partial F}{\partial y}\right)\Big|_{(x, g(x)).}$$

Proof. Let $\frac{\partial F}{\partial x}\big|_{(x_0, y_0)} = \alpha$, and $\frac{\partial F}{\partial y}\Big|_{(x_0, y_0)} = \beta$. We define a mapping $H : D \to \mathbb{R}$ by

$$H(x, y) = y - \frac{F(x, y)}{\beta} \quad \text{for all } (x, y) \text{ in } D.$$

Then $H(x,y)$ is continuous in D, $H(x_0,y_0)=y_0$, and $\left.\frac{\partial H}{\partial y}\right|_{(x_0,y_0)}=0$. Since $\frac{\partial H}{\partial y}$ is continuous in D, there exist real numbers $h>0$ and $k>0$ such that

$$\frac{\partial H}{\partial y}<\frac{1}{2} \quad \text{for all } (x,y) \text{ in } R=\{(x,y)\in\mathbb{R}^2 : |x-x_0|<h \text{ and } |y-y_0|<k\}.$$

Since $H(x,y)$ is continuous, and $H(x_0,y_0)=y_0$, we can take a smaller h if necessary so that $|H(x,y_0)-y_0|<\frac{k}{2}$ for all x in $[x_0-h,x_0+h]$. Now let X be the set of all continuous real valued functions $g:[x_0-h,x_0+h]\to[y_0-k,y_0+k]$ such that $g(x_0)=y_0$. X is obviously non-empty. Let $d:X\times X\to\mathbb{R}$ be a mapping defined by $d(g_1,g_2)=\sup\{|g_1(x)-g_2(x)| : x\in[x_0-h,x_0+h]\}$. Then (X,d) is easily seen to be a complete metric space. Now we define a mapping $T:X\to X$ by $T(g)=h$, where $h(x)=H(x,g(x))$ for all $x\in[x_0-h,x_0+h]$. Then h is continuous in $[x_0-h,x_0+h]$, $h(x_0)=H(x_0,g(x_0))=y_0$, and

$$\begin{aligned}|h(x)-y_0| &= |H(x,g(x))-y_0|\\ &\le |H(x,g(x))-H(x,y_0)|+|H(x,y_0)-y_0|\\ &= \left.\frac{\partial H}{\partial y}\right|_{(x,\xi)}|g(x)-y_0|+\frac{k}{2}\end{aligned}$$

for some ξ between $g(x)$ and y_0 by Lagrange's mean value theorem. Therefore, $|h(x)-y_0|\le k$. Thus, T is indeed a mapping from X into itself. Also

$$\begin{aligned}|h_1(x)-h_2(x)| &= |H(x,g_1(x))-H(x,g_2(x))|\\ &= \left.\frac{\partial H}{\partial y}\right|_{(x,\eta)}|g_1(x)-g_2(x)|\\ &\le \frac{|g_1(x)-g_2(x)|}{2}.\end{aligned}$$

This shows that T is a contraction mapping. Therefore, T has a unique fixed point g in X so that $T(g)=g$, that is, $g(x)=H(x,g(x))$. Thus, $F(x,g(x))=0$ for all x in $[x_0-h,x_0+h]$. Now it remains to prove that g is differentiable. Let x and $x+\Delta x$ be any points in $[x_0-h,x_0+h]$. Then $F(x,g(x))=0=F(x+\Delta x,g(x+\Delta x))$, and

$$\begin{aligned}0 &= F(x+\Delta x,g(x+\Delta x))-F(x,g(x))\\ &= F(x+\Delta x,g(x+\Delta x))-F(x+\Delta x,g(x))+F(x+\Delta x,g(x))-F(x,g(x))\\ &= \left.\frac{\partial F}{\partial y}\right|_{(x+\Delta x,\xi')}(g(x+\Delta x)-g(x))+\left.\frac{\partial F}{\partial x}\right|_{(\eta',g(x))}\Delta x\end{aligned}$$

for some ξ' between $g(x+\Delta x)$ and $g(x)$, and η' between x and $x+\Delta x$. Therefore,

$$\frac{g(x+\Delta x)-g(x)}{\Delta x}=-\left(\frac{\partial F}{\partial x}\right)_{(\eta',g(x))}\Big/\left(\frac{\partial F}{\partial y}\right)_{(x+\Delta x,\xi').}$$

Now, as Δx tends to zero, η' tends to x, and ξ' tends to $g(x)$, and so

$$g'(x)=\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x}=-\frac{\partial F}{\partial x}\Big/\left.\frac{\partial F}{\partial y}\right|_{(x,g(x)).}$$

The inverse function theorem given below follows as a Corollary of the implicit function theorem 9.2.6.

Theorem 9.2.7 Let h_0 be a positive real number, and $f :]x_0 - h_0, x_0 + h_0[\to \mathbb{R}$ be a function such that

(i) $f(x_0) = y_0$,

(ii) $f'(x)$ is continuous in $]x_0 - h_0, x_0 + h_0[$, and

(iii) $f'(x_0) \neq 0$.

Then there exists a rectangle $R = \{(x, y) \in \mathbb{R}^2 : |x - x_0| < h, \text{ and } |y - y_0| < k\}$, where $h < h_0$, and a function $g : [y_0 - k, y_0 + k] \to [x_0 - h, x_0 + h]$ satisfying

1. $f(g(y)) = y$ for all y in $[y_0 - k, y_0 + k]$, and
2. $g'(y) = \frac{1}{f'(x)}$, where $x = g(y)$ for all y in $[y_0 - k, y_0 + k]$.

Proof. Let $D = \{(x, y) \in \mathbb{R}^2 : x \in]x_0 - h_0, x_0 + h_0[, \text{ and } y \text{ is arbitrary}\}$. Then D is an open set in $\mathbb{R}^2$. Let $F : D \to \mathbb{R}$ be a mapping defined by

$$F(x, y) = y - f(x).$$

Then $F(x, y)$, $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ are continuous in D. Also $F(x_0, y_0) = y_0 - f(x_0) = 0$, and $\frac{\partial F}{\partial x}|_{(x_0,y_0)} \neq 0$. Then, by the implicit function theorem (with roles of x and y interchanged), there exist a rectangle $R = \{(x, y) \in \mathbb{R}^2 : |x - x_0| < h, \text{ and } |y - y_0| < k\}$, where $h < h_0$ and a function $g : [y_0 - k, y_0 + k] \to [x_0 - h, x_0 + h]$ such that

1. $F(g(y), y) = 0$ for all y in $[y_0 - k, y_0 + k]$, and
2. g is differentiable in $[y_0 - k, y_0 + k]$, and its derivative

$$g'(y) = -\frac{\partial F}{\partial y} \Big/ \frac{\partial F}{\partial x}\Big|_{(g(y),y).}$$

From (1) it follows that $f(g(y)) = y$ for all y in $[y_0 - k, y_0 + k]$, that is, g is the inverse of f. Also, from (2)

$$g'(y) = \frac{1}{\left(\frac{\partial F}{\partial x}\right)_{(g(y),y)}} = \frac{1}{f'(x)}, \text{ where } x = g(y).$$

As a further application of Banach Theorem 9.1.4, we consider a system of linear equations.

Theorem 9.2.8 Let $A = [a_{ij}]$ be a square matrix of order n of real numbers. Let $x = Ax + B$, where $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, be a system of n linear equations in unknowns $x_1, x_2, \ldots, x_n$. If $\sum_{j=1}^{n} |a_{ij}| < 1$ for all $i = 1, 2, \ldots, n$, then the system of linear equations has a unique solution.

Proof. Let X be the set of n-rowed column vectors of real numbers. Let us define a mapping $d : X \times X \to \mathbb{R}$ by

$$d(x, y) = \max_i \{|x_i - y_i|\}$$

for all $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in X. Then (X, d) is a complete metric space. Let $T : X \to X$ be a maping given by

$$T(x) = Ax + B$$

where A and B are as given in the statement of the theorem. Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ be any points in X. Then

$$T(x) = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix} + B \quad \text{and} \quad T(y) = \begin{bmatrix} \sum_{i=1}^{n} a_{1j}y_j \\ \sum_{j=1}^{n} a_{2j}y_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}y_j \end{bmatrix} + B.$$

Therefore,

$$d(T(x), T(y)) = \max_i \left\{ \left| \sum_{j=1}^{n} a_{ij}(x_j - y_j) \right| \right\}$$

where maximum is taken over all $i = 1, 2, \ldots, n$. Then

$$\begin{aligned} d(T(x), T(y)) &\leq \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \, |x_j - y_j| \right\} \\ &\leq \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \right\} d(x, y) \\ &= \alpha d(x, y). \end{aligned}$$

where $\alpha = \max_i \{\sum_{j=1}^{n} |a_{ij}|\} < 1$ by our assumption. Thus, T is a contraction mapping, and so there is a unique fixed point x in X of T. Then $x = Ax + B$, and the system of linear equations has a unique solution x.

As is the case in Theorem 9.1.4, the unique solution x of the system of linear equations can be iterated from an arbitrary point x_0 in X by the sequence

$$x_1 = Ax_0 + B,$$

$$x_{n+1} = T(x_n) = Ax_n + B \text{ for } n \geq 1.$$

Then $\lim_{n\to\infty} x_n = x$.

Example 9.2.9 We consider the following system of linear equations

$$\left.\begin{aligned} x &= \frac{1}{3}x - \frac{1}{4}y + \frac{1}{4}z - 1 \\ y &= -\frac{1}{2}x + \frac{1}{5}y + \frac{1}{4}z + 2 \\ z &= \frac{1}{5}x - \frac{1}{3}y + \frac{1}{4}z - 2 \end{aligned}\right\}. \tag{9.9}$$

The value of α for this system of linear equation is given by

$$\alpha = \max\left\{\frac{1}{3} + \frac{1}{4} + \frac{1}{4},\ \frac{1}{2} + \frac{1}{5} + \frac{1}{4},\ \frac{1}{5} + \frac{1}{3} + \frac{1}{4}\right\}$$

which is less than 1. So, by the above theorem, the system (9.9) has a unique solution.

EXERCISE

1. Let $f : [0,1] \to [0,1]$ be a function defined by $f(x) = 4x - 4x^2$
   (a) Draw the graph of f
   (b) Is f one-one?
   (c) Find the fixed points of the mapping f.
2. Give an example of a continuous mapping of the open interval $]0,1[$ onto itself having no fixed point. Show that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = (x^2+1)^{\frac{1}{2}}$ has no fixed point.
3. Show that each mapping of a semi-open interval onto itself has at least one fixed point.
4. Give an example of a mapping of $[0,1]$ onto itself having exactly two fixed point, namely 0 and 1.
5. $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = e^{-x}$. Use differentiation to show that f is not a contraction on $\mathbb{R}$, but that f^2 is. Use an iterative technique to find (upto three decimal places) the unique fixed point of f, and hence the unique root of $\log x + x = 0$.
6. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x + e^{\frac{-x}{2}} & \text{if } x \geq 0, \\ e^{\frac{x}{2}} & \text{if } x < 0. \end{cases}$$

   Show that $|f(x) - f(y)| < |x - y|$ for all x and y, yet f has no fixed point. Does this contradict the contraction mapping theorem? Explain.

7. For each of the following sets, give an example of a continuous mapping of the set into itself that has no fixed point: (a) $\mathbb{R}$ (b) $]0,1]$ (c) $[10,-7]\cup[2,4]$.

8. Show that the method of successive approximation applied to the differential equation $f'=f$ with $f(0)=1$ yields the usual formula for e^x.

9. Given an example of a mapping of [0,1] into itself that is not continuous, and has no fixed point.

10. Does the relation $x^3+y^3+x-y=0$ represent a continuous and differentiable function in some neighbourhood of zero?

11. Let $f:\mathbb{R}^2\to\mathbb{R}^2$ be defined by $f(x,y)=(x',y')$, where

$$x'=\frac{\sin x}{3}+\frac{\cos y}{3}+2,$$
$$y'=\frac{\cos x}{6}+\frac{\sin y}{2}-1.$$

Show that f has a fixed point.

12. Show that the following system of linear equations

$$\begin{aligned} x_1 &= \frac{1}{4}x_1-\frac{1}{4}x_2+\frac{2}{15}x_3+3\\ x_2 &= \frac{1}{4}x_1+\frac{1}{5}x_2+\frac{1}{2}x_3-1\\ x_3 &= -\frac{1}{4}x_1+\frac{1}{3}x_2-\frac{1}{3}x_3+2 \end{aligned}$$

has a unique solution using Banach's fixed point theorem.

13. Show that the equation $\frac{dy}{dx}=y^{\frac{1}{3}}$, $y(0)=1$ does not have a unique solution.

[Hint. The solutions $\phi(x)=0$ and $\phi(x)=(\frac{2x}{3})^{\frac{3}{2}}$ of the differential equation does not satisfy Lipschitz condition in any neighbourhood of $(0,0)$.]

14. Find an interval I in which the relation $x^3+y^3-6xy=0$ defines y as a function of $x\in I$. Also find an interval J in which the above relation defines x as a function of $y\in J$.

Chapter 10

A Construction of Real Number System

In Section 1.8, we introduced the integers as equivalence classes of ordered pairs of natural numbers, and the rational numbers as equivalence classes of ordered pairs of integers. In the construction of real numbers, we begin with the set of all Cauchy sequence of rational numbers, and an equivalence relation is defined on it. A real number is then defined to be an equivalence class of a Cauchy sequence of rational numbers. The set of real numbers obtained in this way form an ordered field which is complete as a metric space. The ordered field of rational numbers can be isomorphically embedded in this field of real numbers. The construction of this complete ordered field of real numbers from the incomplete ordered field of rational numbers is analogous to the construction of a completion of an incomplete metric space (see Theorem 5.4.1).

10.1 Ordered Field of Real Numbers $\mathbb{R}_C$

In the following, we consider the ordered field of rational numbers $(\mathbb{Q}, +, \cdot \leq)$ in which the distance is defined by

$$d(x, y) = |x - y| \quad \text{for all } x \text{ and } y \text{ in } \mathbb{Q}.$$

Definition 10.1.1 A sequence $\langle x_n \rangle$ of rational numbers is said to be a Cauchy sequence (or a fundamental sequence) if for each rational number $\epsilon > 0$, there exists a $p \in \mathbb{Z}_+$ such that

$$|x_n - x_m| < \epsilon \quad \text{for all } m, n \geq p.$$

We denote by $F_{\mathbb{Q}}$ the set of all Cauchy sequences of rational numbers.

The notion of convergence of a sequence $\langle x_n \rangle$ in $\mathbb{Q}$ is defined as follows:

Definition 10.1.2 A sequence $\langle x_n \rangle$ of rational numbers is said to converge to a rational number x if for each rational number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that

$$|x_n - x| < \epsilon \quad \text{for all } n \geq m.$$

As usual, if $\langle x_n \rangle$ in $\mathbb{Q}$ converges to $x \in \mathbb{Q}$, we write $\langle x_n \rangle \to x$, or $\lim_{n\to\infty} x_n = x$.

Definition 10.1.3 A sequence $\langle x_n \rangle$ of rational numbers is said to be bounded if there exists a rational number k such that $|x_n| \leq k$ for all n.

It is easily seen that every Cauchy sequence $\langle x_n \rangle$ in $\mathbb{Q}$ is bounded. Also, it is noted that if $\langle x_n \rangle$ and $\langle y_n \rangle$ are Cauchy sequences of rational numbers, then $\langle x_n + y_n \rangle$, $\langle x_n - y_n \rangle$ and $\langle x_n y_n \rangle$ are Cauchy sequences of rational numbers; further, if $\langle y_n \rangle$ in $F_{\mathbb{Q}}$ is such that $y_n \neq 0$ for all n, and $\langle y_n \rangle$ does not converge to zero in $\mathbb{Q}$, then $\left\langle \frac{1}{y_n} \right\rangle$ is in $F_{\mathbb{Q}}$, and thus $\left\langle \frac{x_n}{y_n} \right\rangle$ is a Cauchy sequence of rational numbers.

In $F_{\mathbb{Q}}$, we define a relation $\sim$ as follows:

$$\langle x_n \rangle \sim \langle y_n \rangle \text{ if } \lim_{n \to \infty} (x_n - y_n) = 0.$$

Note that $x_n - y_n = x_n + (-y_n)$ for all n. Clearly, the relation $\sim$ is an equivalence relation in $F_{\mathbb{Q}}$, and therefore it partitions $F_{\mathbb{Q}}$ into equivalence classes. We denote the equivalence class of $\langle x_n \rangle$ in $F_{\mathbb{Q}}$ by $[\langle x_n \rangle]$, that is,

$$[\langle x_n \rangle] = \{\langle y_n \rangle \in F_{\mathbb{Q}} : \langle y_n \rangle \sim \langle x_n \rangle\}.$$

Definition 10.1.4 A real number is an equivalence class $[\langle x_n \rangle]$ with respect to the relation $\sim$, where $\langle x_n \rangle$ is a Cauchy sequence of rational numbers. We denote by $\mathbb{R}_{\mathbb{C}}$ the set of all real numbers, that is,

$$\mathbb{R}_{\mathbb{C}} = \{[\langle x_n \rangle] : \langle x_n \rangle \in F_{\mathbb{Q}}\}.$$

Since two equivalence classes of an equivalence relation are either disjoint or equal, it follows that two real numbers $[\langle x_n \rangle]$ and $[\langle y_n \rangle]$ are equal written $[\langle x_n \rangle] = [\langle y_n \rangle]$, if and only if $\langle x_n \rangle \sim \langle y_n \rangle$.

We now define the two binary algebraic operations in $\mathbb{R}_C$ called addition and multiplication.

Definition 10.1.5 Let $\xi = [\langle x_n \rangle]$ and $\eta = [\langle y_n \rangle]$ be two real numbers in $\mathbb{R}_{\mathbb{C}}$. Then the sum of ξ and η, denoted $\xi \oplus \eta$, is defined as

$$\xi \oplus \eta = [\langle x_n + y_n \rangle].$$

To show that the above definition of addition is well-defined, we need to show that if we take sequences $\langle x'_n \rangle$ in ξ and $\langle y'_n \rangle$ in η, then $[\langle x_n + y_n \rangle] = [\langle x'_n + y'_n \rangle]$, that is, the addition $\xi \oplus \eta$ does not depend on which representative sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ are chosen from ξ and η, respectively. So we assume that $\langle x'_n \rangle \in \xi$ and $\langle y'_n \rangle \in \eta$. Then

$$\lim_{n \to \infty} [x_n + y_n - (x'_n + y'_n)] = \lim_{n \to \infty} (x_n - x'_n) + \lim_{n \to \infty} (y_n - y'_n) = 0.$$

Thus, $\langle x_n + y_n \rangle \sim \langle x'_n + y'_n \rangle$ so that $[\langle x_n + y_n \rangle] = [\langle x'_n + y'_n \rangle]$.

Next the product is defined.

Definition 10.1.6 Let $\xi = [\langle x_n \rangle]$ and $\eta = [\langle y_n \rangle]$ be two real numbers. The product of ξ and η, denoted $\xi \odot \eta$, is defined as

$$\xi \odot \eta = [\langle x_n y_n \rangle].$$

Again we need to show that the above product is well-defined, that is, if we choose $\langle x'_n \rangle$ and $\langle y'_n \rangle$ in ξ and η, respectively, then

$$[\langle x_n y_n \rangle] = [\langle x'_n y'_n \rangle].$$

But this follows from the inequality

$$|x_n y_n - x'_n y'_n| \leq |x_n|\,|y_n - y'_n| + |y'_n|\,|x_n - x'_n|$$

and the facts that $\langle x_n \rangle$ and $\langle y'_n \rangle$ are bounded seqeunces, and $\lim_{n\to\infty}(x_n - x'_n) = 0$ and $\lim_{n\to\infty}(y_n - y'_n) = 0$.

Theorem 10.1.7 $(\mathbb{R}_C, \oplus, \odot)$ is a field.

Proof. We leave it to the reader to verify the associativity, commutatively and distributivity. The additive identity is $\bar{0} = [\langle 0 \rangle]$, where $\langle 0 \rangle = \langle 0, 0, \dots \rangle$ is the sequence consisting of zeros. The multiplicative identity is $\bar{1} = [\langle 1 \rangle]$, where $\langle 1 \rangle = \langle 1, 1, \dots, \rangle$, the sequence consisting of multiplicative identity 1 in $\mathbb{Q}$. The multiplicative inverse of a non-zero element $\xi = [\langle x_n \rangle]$, that is, $\xi \neq \bar{0}$, can be found in $\mathbb{R}_C$ as follows:

Since $\xi \neq \bar{0}$, x_n can be 0 only for a finite number of n's. Thus, there exists a $q = \mathbb{Z}_+$ such that $x_n \neq 0$ for $n > q$. We define

$$y_n = \begin{cases} 1 & \text{if } n \leq q, \\ \dfrac{1}{x_n} & \text{if } n > q. \end{cases}$$

Then $\xi^{-1} = [\langle y_n \rangle]$ is the multiplicative inverse of ξ.

Next we introduce an order in $\mathbb{R}_C$.

Definition 10.1.8 A sequence $\langle x_n \rangle$ of rational numbers is said to be a positive sequence if there exists a rational number $c > 0$ and a positive integer m such that $x_n \geq c$ for all $n \geq m$.

For example, the sequence $\left\langle \frac{1}{n} \right\rangle$ in $F_{\mathbb{Q}}$ is not a positive sequence. The sum sequence $\langle x_n + y_n \rangle$ and the product sequence $\langle x_n y_n \rangle$ of positive sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ of rational numbers are positive sequences. Also, if $\xi = [\langle x_n \rangle]$ is a real number, and $\langle x_n \rangle$ is a positive sequence, then every Cauchy sequence $\langle x'_n \rangle$ in ξ is also a positive sequence since $\langle x'_n \rangle \sim \langle x_n \rangle$. This can be proved as follows:

Since $\langle x_n \rangle$ is a positive sequence, there exists a rational number $c > 0$ and a positive integer n_1 such that $x_n \geq c$ for all $n \geq n_1$. Since $\langle x_n \rangle \sim \langle x'_n \rangle$, there exists a positive integer n_2 such that

$$|x_n - x'_n| < \frac{c}{2} \quad \text{for all } n \geq n_2.$$

Then $x'_n = (x'_n - x_n) + x_n > \frac{-c}{2} + c = \frac{c}{2}$ for all $n \geq \max\{n_1, n_2\}$.

Definition 10.1.9 A real number $\xi = [\langle x_n \rangle]$ is said to be positive if $\langle x_n \rangle$ is a positive sequence in $F_{\mathbb{Q}}$.

By our above remarks, the above definition of a positive real number is well-defined since any other sequence in $\xi = [\langle x_n \rangle]$ is positive if $\langle x_n \rangle$ is positive. We denote by $\mathbb{R}_{C+}$ the set of positive real numbers.

It follows that if ξ and η are in $\mathbb{R}_{C+}$, then their sum $\xi \oplus \eta$ and product $\xi \odot \eta$ are also in $\mathbb{R}_{C+}$. Also, if $\xi = [\langle x_n \rangle] \in \mathbb{R}_c$, then $-\xi$ denotes the real number $[\langle -x_n \rangle]$.

Theorem 10.1.10 If $\xi \in \mathbb{R}_c$, then one and only one of the following holds:
(i) $\xi \in \mathbb{R}_{C+}$ (ii) $\xi = \bar{0}$ (iii) $-\xi \in \mathbb{R}_{C+}$

Proof. First we show that at least one of the three possibilities holds. Let $\xi = [\langle x_n \rangle]$. If $\xi \neq \bar{0}$, then $\langle x_n \rangle$ is not in $\bar{0} = [\langle 0 \rangle]$, that is, $\lim_{n\to\infty} x_n \neq 0$. This means that there is a

rational number $\epsilon > 0$ such that for each positive integer m, there exists a positive integer $n > m$ such that

$$|x_n - 0| > \epsilon. \tag{10.1}$$

Since $\langle x_n \rangle$ is a Cauchy sequence, there exists a positive integer p such that

$$|x_n - x_m| < \frac{\epsilon}{2} \quad \text{for all } n, m \geq p.$$

By statement (10.1), there exists an $n > p$ such that $|x_n| > \epsilon$. Therefore,

$$x_n - \frac{\epsilon}{2} < x_m < x_n + \frac{\epsilon}{2} \quad \text{for all } m \geq p.$$

Two cases arise: $x_n > \epsilon$, or $x_n < -\epsilon$. In the first case, $x_m > \frac{\epsilon}{2}$ for all $m \geq p$ so that $\xi \in \mathbb{R}_{C+}$. In the later case, $x_m < \frac{-\epsilon}{2}$ for all $m \geq p$ which means that $-\xi \in \mathbb{R}_{C+}$.

We now prove that at most one of the thee alternatives is true. If $\xi = \bar{0}$, then $\lim_{n\to\infty} x_n = 0$. Therefore, for each rational number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that $|x_n| < \epsilon$ for all $n \geq m$. Hence, neither $\langle x_n \rangle$ is a positive sequence nor $\langle -x_n \rangle$ is a positive sequence. This means that $\xi \notin \mathbb{R}_{C+}$ and $-\xi \notin \mathbb{R}_{C+}$. If $\xi \neq \bar{0}$, then we show that it is not possible that $\xi \in \mathbb{R}_{C+}$, and $-\xi \in \mathbb{R}_{C+}$. Suppose on the contrary that ξ and $-\xi$ are both in $\mathbb{R}_{C+}$. Then both $\langle x_n \rangle$ and $\langle -x_n \rangle$ are positive sequences. Then there exist rational numbers $\epsilon > 0$ and $\epsilon' > 0$ and positive integers m and m' such that

$$x_n \geq \epsilon \quad \text{for all } n \geq m, \text{ and } -x_n \geq \epsilon' \quad \text{for all } n \geq m'.$$

If $n > \max\{m, m'\}$, then

$$0 < \epsilon \leq x_n \leq -\epsilon' < 0$$

which is a contradiction. This completes the proof.

Now we are in a position to introduce the following order in $\mathbb{R}_C$.

Definition 10.1.11 Let ξ and η be two real numbers. ξ is said to be greater than η, written $\xi > \eta$, if $\xi \ominus \eta \in \mathbb{R}_{C+}$, where $\xi \ominus \eta = \xi \oplus (-\eta)$.

As usual we write $\xi \geq \eta$ if either $\xi > \eta$ or $\xi = \eta$. Similarly, $\xi < \eta$ if $\eta > \xi$, and $\xi \leq \eta$ if $\xi < \eta$ or $\xi = \eta$.

Definition 10.1.12 The modulus of a real number ξ (or the absolute value of ξ), denoted by $|\xi|$, is defined as

$$|\xi| = \begin{cases} \xi & \text{if } \xi \geq \bar{0}, \\ -\xi & \text{if } \xi < \bar{0}. \end{cases}$$

The metric on $\mathbb{R}_C$ is defined by

$$d(\xi, \eta) = |\xi - \eta| \quad \text{for all } \xi, \eta \in \mathbb{R}_C.$$

That the mapping d is actually a metric follows from the following properties of the modulus function.

1. $|\xi| \geq \bar{0}$,
2. $|\xi| = \bar{0}$ if and only if $\xi = \bar{0}$,

3. $|\xi| = |-\xi|$,
4. $|\xi \oplus \eta| \leq |\xi| \oplus |\eta|$.

The following theorem is left to the reader as an exercise.

Theorem 10.1.13 Let ξ, η and ζ be real numbers in $\mathbb{R}_C$. The following statements are true.

1. The relation $\leq$ is a partial order on $\mathbb{R}_C$.
2. The law of Trichotomy: one and only one of the following three conditions holds
   (a) $\xi > \eta$ (b) $\xi = \eta$ (c) $\eta > \xi$.
3. If $\xi \leq \eta$, then $\xi \oplus \zeta \leq \eta \oplus \zeta$.
4. If $\xi \leq \eta$, and $\zeta \geq \bar{0}$, then $\xi \odot \zeta \leq \eta \odot \zeta$.

By the above theorem, it follows that $(\mathbb{R}_C, +, \cdot, \leq)$ is an ordered field.

Theorem 10.1.14 The order relation $\leq$ on $\mathbb{R}_C$ is dense, that is, if ξ and η are in $\mathbb{R}_C$ and $\xi < \eta$, then there exists a $\zeta \in \mathbb{R}_C$ such that $\xi < \zeta < \eta$.

Proof. If $\xi < \eta$, then

$$2 \odot \xi = \xi \oplus \xi < \xi \oplus \eta < \eta \oplus \eta = 2 \odot \eta.$$

Since $2 = \bar{1} \oplus \bar{1} > \bar{0}$, it follows that the multiplicative inverse of 2, denoted by $\frac{1}{2}$, is greater than zero, that is, $\frac{1}{2} > \bar{0}$, and $\xi < \frac{1}{2} \odot (\xi \oplus \eta) < \eta$.

EXERCISE

1. Let $\langle x_n \rangle$ be a Cauchy sequence in $\mathbb{Q}$. Then exactly one of the following statements holds:
   (a) $\lim_{n\to\infty} x_n = 0$.
   (b) $\langle x_n \rangle$ is a positive sequence.
   (c) $\langle -x_n \rangle$ is a positive sequence.
2. Show that a Cauchy sequence $\langle x_n \rangle$ is positive if and only if $\langle |x_n| \rangle$ is positive.
3. Let $\langle x_n \rangle$ be a convergent sequence in an ordered field A, then $\langle x_n \rangle$ is positive if and only if $\lim_{n\to\infty} x_n > 0$ in A.
4. Show that the sequence

$$x_1 = 1,$$
$$x_{n+1} = x_1 + \frac{x_2}{2} + \frac{x_3}{3} + \cdots + \frac{x_{n-1}}{n-1}$$

   is a Cauchy sequence of rational numbers.

5. Let $\langle x_n \rangle$ be a sequence in $F_{\mathbb{Q}}$ such that for any rational number $\epsilon > 0$ and $m \in \mathbb{Z}_+$, there exists an $n \in \mathbb{Z}_+$ such that $n > m$, and $|x_n| < \epsilon$, then $\langle x_n \rangle \in \bar{0}$.

6. Let ξ, η and ζ be real numbers. Prove the following:

   (a) If $\xi > \eta$ and $-\zeta \in \mathbb{R}_{C+}$, then $\xi \odot \zeta < \eta \odot \zeta$.
   (b) If $\xi \in \mathbb{R}_{C+}$, then $\xi^{-1} \in \mathbb{R}_{C+}$.
   (c) $\bar{0} \leq \xi^2$, and $\bar{0} = \xi^2$ if and only if $\xi = \bar{0}$.

7. Let ξ and η be real numbers. Prove the following statements.

   (a) $|\xi| = \max\{\xi, -\xi\}$, $|\xi|^2 = \xi^2 = |-\xi|^2$, $|-\xi| = |\xi|$.
   (b) $|\xi \oplus \eta| \leq |\xi| \oplus |\eta|$.
   (c) $|\zeta \odot \eta| = |\zeta| \odot |\eta|$.
   (d) $|\xi - \eta| \geq \|\xi| - |\eta\| \geq |\xi| - |\eta|$.
   (e) If $\langle x_n \rangle \in \xi$, then $|\xi| = [\langle |x_n| \rangle]$.

10.2 Embedding of Rational Numbers in $\mathbb{R}_C$

Theorem 10.2.1 (Embedding) The mapping $f : \mathbb{Q} \to \mathbb{R}_C$ defined by $f(x) = [\langle x \rangle]$, where $\langle x \rangle$ is the constant sequence $\langle x, x, \dots \rangle$, is an isomorphism of $\mathbb{Q}$ into $\mathbb{R}_C$ preserving addition, multiplication and order.

Proof. f is injective. For $[\langle x \rangle] = [\langle y \rangle]$ if and only if $\langle x \rangle \sim \langle y \rangle$ if and only if $x = y$ in $\mathbb{Q}$. Now we prove the following:

$$f(x + y) = f(x) \oplus f(y),$$
$$f(xy) = f(x) \odot f(y)$$

and $x < y$ implies that $f(x) < f(y)$ for all x and y in $\mathbb{Q}$.

If $x, y \in \mathbb{Q}$, then

$$f(x + y) = [\langle x + y \rangle] = [\langle x \rangle] \oplus [\langle y \rangle] = f(x) \oplus f(y),$$

and

$$f(xy) = [\langle xy \rangle] = [\langle x \rangle] \odot [\langle y \rangle] = f(x) \odot f(y).$$

Thus, f preserves addition and multiplication.

Now $x < y$ in Q if and only if $y - x > 0$ in $\mathbb{Q}$ if and only if the sequence $\langle y - x \rangle = \langle y - x, y - x, \dots \rangle$ is a positive sequence of rational numbers if and only if $[\langle y - x \rangle] \in \mathbb{R}_{C+}$ if and only if $[\langle y \rangle] \oplus [\langle -x \rangle] \in \mathbb{R}_{C+}$ if and only if $[\langle y \rangle] \ominus [\langle x \rangle] \in \mathbb{R}_{C+}$ if and only if $f(x) < f(y)$. This completes the proof.

As usual, we identify $\mathbb{Q}$ with its isomorphic image $f(\mathbb{Q})$ in $\mathbb{R}_C$, and use interchangeably the symbols x and $[\langle x \rangle]$. Due to this isomrphism, $[\langle x \rangle]$ is sometimes called a real rational number for each x in $\mathbb{Q}$. A real number which is not a rational number is called an irrational number. Also, when there is no confusion, the symbols $\oplus, \odot$ are replaced by the symbols

$+, \cdot$, respectively, $\bar{0}$ is denoted by 0, and $\bar{1}$ is written as 1. Already, we are using the same symbols $<, \leq$ for the order in $\mathbb{Q}$ and $\mathbb{R}_C$.

EXERCISE

1. Show that the sequence defined by

$$x_1 = 1,$$
$$x_{n+1} = x_n + \frac{a_n}{10^n}, \quad n \geq 1,$$

where a_n is a non-negative integer such that

$$\left(x_n + \frac{a_n}{10^n}\right)^2 < 2 < \left[x_n + \frac{a_n + 1}{10^n}\right]^2,$$

is a Cauchy sequence but not convergent in $\mathbb{Q}$. (The x_n's are decimal approximations to $\sqrt{2}$.)

10.3 Completeness of $\mathbb{R}_C$

We show that $\mathbb{R}_C$ is a complete metric space with respect to the metric $d(\xi, \eta) = |\xi - \eta|$ for all ξ and η in $\mathbb{R}_C$. For this we need to define the notion of convergence of a sequence in $\mathbb{R}_C$, and the notion of being a Cauchy sequence in $\mathbb{R}_C$.

Definition 10.3.1 A sequence $\langle x_n \rangle$ in $\mathbb{R}_C$ converges to a point $x \in \mathbb{R}_C$ if for each real number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that

$$|x_n - x| < \epsilon \quad \text{for all } n \geq m.$$

Definition 10.3.2 A sequence $\langle x_n \rangle$ in $\mathbb{R}_C$ is said to be a Cauchy sequence if for each real number $\epsilon > 0$, there exists a $p \in \mathbb{Z}_+$ such that

$$|x_n - x_m| < \epsilon \quad \text{for all } m, n \geq p.$$

It follows immediately that a convergent sequence in $\mathbb{R}_C$ is a Cauchy sequence. The converse of this statement is also true; every Cauchy sequence in $\mathbb{R}_C$ is convergent. Before proving this, we prove that every Cauchy sequence of rational numbers converges in $\mathbb{R}_C$.

Theorem 10.3.3 Let $\xi \in \mathbb{R}_C$, and $\langle x_n \rangle \in \xi$. Then $\lim_{n \to \infty} x_n = \xi$ in $\mathbb{R}_C$.

Proof. Let $\epsilon > 0$ be a real number in $\mathbb{R}_C$. Then there exists an ϵ' in $\mathbb{Q}$ such that $0 < \epsilon' < \epsilon$. Since $\langle x_n \rangle$ is a Cauchy sequence in $\mathbb{Q}$, there exists an $m \in \mathbb{Z}_+$ such that

$$|x_p - x_q| < \frac{\epsilon'}{2} \quad \text{for all } p, q \geq m.$$

Let p be a fixed positive integer, and define

$$y_q = \epsilon' - |x_p - x_q| \quad \text{for all } q \in \mathbb{Z}_+.$$

Then for each $p \geq m$, the sequence $\langle y_q \rangle$ is a positive sequence since

$$y_q = \epsilon' - |x_p - x_q| > \frac{\epsilon'}{2} \quad \text{for all } q \geq m.$$

Let $\eta = [\langle y_q \rangle]$. Then $\eta \in \mathbb{R}_{C+}$ so that

$$[\langle y_q \rangle] = \epsilon' - [\langle |x_p - x_q| \rangle] > 0 \quad \text{in } \mathbb{R}_C.$$

Thus, for $p \geq m$

$$\begin{aligned} |x_p - \xi| &= |x_p - [\langle x_q \rangle]| \\ &= |[\langle x_p \rangle] - [\langle x_q \rangle]| \\ &= |[\langle x_p - x_q \rangle]| \\ &= [\langle |x_p - x_q| \rangle] \\ &< \epsilon' < \epsilon. \end{aligned}$$

Therefore, $\lim_{n\to\infty} x_n = \xi$.

Corollary 10.3.4 If $\xi \in \mathbb{R}_C$, and $\epsilon > 0$ be a real number, then there exists a rational number x such that $|\xi - x| < \epsilon$ in $\mathbb{R}_C$, that is, rational numbers are dense in $\mathbb{R}_C$.

Proof. Let $\langle x_n \rangle \in \xi$. Then $\lim_{n\to\infty} x_n = \xi$. Then for real number $\epsilon > 0$, there exists an $m \in \mathbb{Z}_+$ such that

$$|x_n - \xi| < \epsilon \quad \text{for all } n \geq m.$$

In particular, $|x_m - \xi| < \epsilon$ as required.

Corollary 10.3.5 If $\xi < \eta$ in $\mathbb{R}_C$, then there exists a rational number z such that $\xi < z < \eta$.

Proof. By Theorem 10.1.14, there exists a real number ζ such that $\xi < \zeta < \eta$. Let $\epsilon = \min\{\zeta - \xi, \eta - \zeta\}$. Then, by Corollary 10.3.4, there exists a rational number z such that

$$|\zeta - z| < \epsilon.$$

Then $\xi < \zeta - \epsilon < z < \zeta + \epsilon < \eta$.

Corollary 10.3.6 $\mathbb{R}_C$ is Archimedean, that is, if ξ and $\eta \in \mathbb{R}_{C+}$, then there exists an $n \in \mathbb{Z}_+$ such that $n\xi > \eta$.

Proof. Let $0 < \xi < \eta$ in $\mathbb{R}_C$. Let x and y be rational numbers such that $0 < x < \xi < \eta \leq y < \xi + \eta$. Since $\mathbb{Q}$ is Archimedean, and embedding isomorphism preserves addition and order, there exists an $n \in \mathbb{Z}_+$ such that $nx \geq y$ in $\mathbb{R}_C$. But then

$$n\xi > nx \geq y \geq \eta.$$

Thus, $\mathbb{R}_C$ is Archimedean.

For an example of an ordered field which is not Archimedean, the reader is referred to Cohen and Ehrlich [11].

Theorem 10.3.7 The metric space $(\mathbb{R}_C, d)$, where $d(\xi, \eta) = |\xi - \eta|$ for all $\xi, \eta \in \mathbb{R}_C$, is complete.

Proof. Let $\langle x_n \rangle$ be a Cauchy sequence in $\mathbb{R}_C$, and $\epsilon > 0$ be a real number. By Corollary 10.3.4, for each $p \in \mathbb{Z}_+$, there exists a rational number a_p such that

$$|x_p - a_p| < \frac{1}{p}.$$

Since $\langle x_n \rangle$ is a Cauchy sequence, there exists an $m_1 \in \mathbb{Z}_+$ such that

$$|x_p - x_q| < \frac{\epsilon}{3} \quad \text{for all } p, q \geq m_1.$$

Since $\mathbb{R}_C$ is Archimedean, there exists an $m_2 \in \mathbb{Z}_+$ such that $\frac{1}{n} < \frac{\epsilon}{3}$ for all $n \geq m_2$. Then

$$\begin{aligned} |a_p - a_q| &\leq |a_p - x_p| + |x_p - x_q| + |x_q - a_q| \\ &< \epsilon \quad \text{for all } p, q \geq \max\{m_1, m_2\}. \end{aligned}$$

Thus, $\langle a_n \rangle$ is a Cauchy sequence of rational numbers. Let $\xi = [\langle a_n \rangle]$. Then $\lim_{n\to\infty} a_n = \xi$. Therefore, there exists an $m_3 \in \mathbb{Z}_+$ such that

$$|a_n - \xi| < \frac{2\epsilon}{3} \quad \text{for all } n \geq m_3.$$

Then

$$\begin{aligned} |x_n - \xi| &\leq |x_n - a_n| + |a_n - \xi| \\ &< \epsilon \quad \text{for all } n \geq \max\{m_2, m_3\}. \end{aligned}$$

Once the order $\leq$ is defined on $\mathbb{R}_C$, we can define the concepts of an upper bound, lower bound, least upper bound (supremum), greatest lower bound (infimum) of a subset A of $\mathbb{R}_C$ as in Section 1.7.

Theorem 10.3.8 Every non-empty subset of $\mathbb{R}_C$ which has an upper bound has a supremum in $\mathbb{R}_C$.

Proof. Let A be a non-empty subset of $\mathbb{R}_C$, and $b \in \mathbb{R}_C$ be an upper bound A. Let $a \in A$. Then $a \leq b$. Since $\mathbb{R}_C$ is Archimedean, for each $n \in \mathbb{Z}_+$, there exists an $m \in \mathbb{Z}_+$ such that $a + \frac{m}{n} \geq b$. So that $a + \frac{m}{n}$ is an upper bound of A. Let us define

$$K = \{m \in \mathbb{Z}_+ : a + \tfrac{m}{n} \quad \text{is an upper bound of } A\}.$$

Then, by the well-ordering property of $\mathbb{Z}_+$, K has the least element, say, m_n. Then $y_n = a + \frac{m_n}{n}$ is an upper bound of A, and $x_n = a + \frac{(m_n - 1)}{n}$ is not an upper bound of A so that there exists an element $x \in A$ such that $x_n < x$. Therefore, for any $p, q \in \mathbb{Z}_+$

$$x_p < x < y_q \quad \text{for some } x \in A.$$

Then $x_p - x_q < y_q - x_q = \frac{1}{q}$. Similarly, $x_q - x_p < \frac{1}{p}$. Therefore,

$$|x_p - x_q| < \frac{1}{n} \quad \text{for all } p, q \geq n.$$

This shows that $\langle x_n \rangle$ is a Cauchy sequence. Since $\mathbb{R}_C$ is complete, there exists a real number ξ in $\mathbb{R}_C$ such that $\langle x_n \rangle$ converges to ξ. We show that $\xi = \sup A$. First assume that ξ is not an upperbound of A. Then there is a point $x \in A$ such that $\xi < x$. Since $\lim_{n\to\infty} x_n = \xi$, and $\lim_{n\to\infty} \frac{1}{n} = 0$, there exists an $n \in \mathbb{Z}_+$, such that

$$x_n - \xi \leq |x_n - \xi| < \frac{(x-\xi)}{2} \quad \text{and } \frac{1}{n} < \frac{(x-\xi)}{2}.$$

Then $y_n = x_n + \frac{1}{n} < x$, a contradiction since y_n is an upper bound of A. Therefore, ξ is an upper bound of A. Let η be an upper bound of A such that $\eta < \xi$. Then $\xi - \eta > 0$ in $\mathbb{R}_C$. Since $\lim_{n\to\infty} x_n = \xi$, there exists an $n \in \mathbb{Z}_+$ such that $\xi - x_n \leq |\xi - x_n| < \xi - \eta$. Then $\eta < x_n < x$ for some x in A, a contradiction. Hence, $\sup A = \xi$.

Note that to establish the above theorem, we have used the completeness of $\mathbb{R}_C$. It is also possible to take the alternative route in which we first establish theorem 10.3.8, and then prove that $\mathbb{R}_C$ is a complete metric space. The interested reader is referred to Burril [8].

We have proved the following theorem.

Theorem 10.3.9 $(\mathbb{R}_C, +, \cdot, \leq)$ is a complete ordered field.

EXERCISE

1. A sequence in $\mathbb{Q}$ is a Cauchy sequence in $\mathbb{Q}$ if and only if it is a Cauchy sequence in $\mathbb{R}_C$.
2. Let $\langle x_n \rangle$ be a sequence in $\mathbb{Q}$, and $x \in \mathbb{Q}$. Then $\lim_{n\to\infty} x_n = x$ in $\mathbb{Q}$ if and only if $\lim_{n\to\infty} x_n = x$ in $\mathbb{R}_C$.

10.4 Uniqueness of Complete Ordered Field

Between any two complete ordered fields, there is a bijective mapping that preserves addition, multiplication and order, called an isomorphism. First we prove the following lemma.

Definition 10.4.1 Let $\mathbb{F}$ be a complete ordered field. Then the smallest subfield of $\mathbb{F}$ is called the rational subfield of $\mathbb{F}$, denoted by $\mathbb{Q}_\mathbb{F}$.

Since the arbitrary intersection of subfields of a field is a subfield, the above definition makes sense.

Lemma 10.4.2 Let $(\mathbb{F}, +, \cdot, \geq)$ be a complete ordered field. Then for each $a \in \mathbb{F}$, $a = \sup\{r \in \mathbb{Q}_\mathbb{F} : r \leq a\}$.

Proof. Define

$$S_a = \{r \in \mathbb{Q}_\mathbb{F} : r \leq a\}.$$

Since $\mathbb{F}$ is a complete ordered field, it is Archimedean so that there exists an element $r \in \mathbb{Q}_{\mathbb{F}}$ such that $r \leq a$. This shows that S_a is non-empty. Since a is an upper bound of S_a, $\sup S_a$ exists. Let $a' = \sup S_a$. We show that $a = a'$. Two cases arise:

Case (i). $a \in \mathbb{Q}_{\mathbb{F}}$. Then, by the definition, $\sup S_a = a$ so that $a = a'$.

Case (ii). $a \notin \mathbb{Q}_{\mathbb{F}}$. If $a < a'$, then there exists a $p \in \mathbb{Q}_{\mathbb{F}}$ such that $a < p < a'$. Then p is an upper bound of S_a, and $p < a' = \sup S_a$, a contradiction. If $a' < a$, then there exists a $q \in \mathbb{Q}_{\mathbb{F}}$ such that $a' < q < a$. Since $q < a$, $q \in S_a$. Then $q \leq \sup S_a = a' < q$, a contradiction. Thus, $a = a'$ by the law of trichotomy.

The following theorem proves that there is essentially only one complete ordered field.

Theorem 10.4.3 Any two complete ordered fields are isomorphic.

Proof. Let $(\mathbb{F}, +, \cdot, \geq)$ and $(\mathbb{F}', \oplus, \odot, \geq')$ be two complete ordered fields having $\mathbb{Q}_{\mathbb{F}}$ and $\mathbb{Q}_{\mathbb{F}'}$ as their rational subfields, respectively.

Step-1. Let $f : \mathbb{Q} \to \mathbb{F}$ be a mapping defined by

$$\begin{aligned}
&f(1) = e, && \text{where } e \text{ is the identity element of } \mathbb{F},\\
&f(n+1) = f(n) + e && \text{for all } n \in \mathbb{Z}_+,\\
&f(-x) = -f(x) && \text{if } x \in \mathbb{Z}_+,\\
&f(0) = 0, && \text{where 0 in the right is the zero element of } \mathbb{F},\\
&f(\tfrac{p}{q}) = f(p) \cdot [f(q)]^{-1} && \text{for each rational number } \tfrac{p}{q},\ p,\ q \text{ being integers, and } q \neq 0.
\end{aligned}$$

f is clearly injective, and preserves addition, multiplication and order, and $f(\mathbb{Q}) = \mathbb{Q}_{\mathbb{F}}$. Similarly, we define $f' : \mathbb{Q} \to \mathbb{F}'$ so that $f'(\mathbb{Q}) = \mathbb{Q}_{\mathbb{F}'}$. Thus, for each $x \in \mathbb{Q}_{\mathbb{F}}$, there exists a $r \in \mathbb{Q}$ such that $f(r) = x$. We define a mapping $g : \mathbb{Q}_{\mathbb{F}} \to \mathbb{Q}_{\mathbb{F}'}$ such that $g(x) = f'(r)$, where $f(r) = x$, for all $x \in \mathbb{Q}_{\mathbb{F}}$. Since f is a bijection from $\mathbb{Q}$ to $\mathbb{Q}_{\mathbb{F}}$, g is well-defined, and it is a bijection. Also, since f and f' preserves addition, multiplication and order, g also preserves addition, multiplication and order.

Step-2. We show that a subset S of $\mathbb{Q}_{\mathbb{F}}$ is bounded from above if and only if $g(S)$ is bounded from above subset of $\mathbb{Q}_{\mathbb{F}'}$. Let S be a subset of $\mathbb{Q}_{\mathbb{F}}$ bounded from above. Then there exists a $p \in \mathbb{F}$ such that $r \leq p$ for all $r \in S$. Since F is Archimedean, there exists an $n \in \mathbb{Q}_{\mathbb{F}}$ such that $p < n$. Therefore, $r < n$ for all r in S. Thus, $g(r) <' g(n)$ for all r in S. This means that $g(S)$ is bounded from above. Conversely, let S be a subset of $\mathbb{Q}_{\mathbb{F}}$ such that $g(S)$ is bounded from above. Then there exists an element p' in $\mathbb{F}'$ such that $g(r) \leq p'$ for all r in S. Since $\mathbb{F}'$ is Archimedean, there exists an n' in $\mathbb{Q}_{\mathbb{F}'}$ such that $p' <' n'$. If $n \in \mathbb{Q}_{\mathbb{F}}$ is such that $g(n) = n'$, then $g(r) <' g(n)$ for all r in S, that is, $r < n$ for all r in S. Thus, S is bounded from above.

Step-3. For any element $a \in \mathbb{F}$, define

$$S_a = \{r \in \mathbb{Q}_{\mathbb{F}} : r \leq a\}.$$

By Lemma 10.4.2, S_a is bounded from above, and by Step-2, $g(S_a)$ is bounded from above. Consequently, by order completeness of $\mathbb{F}$, $\sup g(S_a)$ exists in $\mathbb{F}'$.

Step-4. Define the mapping $h : \mathbb{F} \to \mathbb{F}'$ by setting $h(a) = \sup g(S_a)$ for all $a \in \mathbb{F}$. We show that h is bijective, and preserves addition multiplication and order. Since for each $a \in \mathbb{F}$, S_a is uniquely defined, and $\sup g(S_a)$ is also unique, therefore, h is well-defined.

Now let a and b be two elements in $\mathbb{F}$ such that $a < b$. Then there exist s and t in $\mathbb{Q}_{\mathbb{F}}$ such that $a < s < t < b$. Then it follows that

$$\sup g(S_a) \leq' g(s) <' g(t) \leq' \sup g(S_b).$$

Thus, $h(a) <' h(b)$ so that h is injective.

To show that h is surjective, let a' be any element in $\mathbb{F}'$. Define

$$S'_{a'} = \{r' \in \mathbb{Q}_{\mathbb{F}'} : r' \leq a'\}$$

and

$$A = \{r \in \mathbb{Q}_{\mathbb{F}} : g(r) \leq' a'\}.$$

Since $S'_{a'}$ is bounded from above, and $g(A) = S'_{a'}$, A is bounded from above by Step-2. Therefore, $\sup A$ exists in $\mathbb{F}$. Let $a = \sup A$. Since $r \in A$ implies that $r \leq a$, $A \subseteq S_a$. Now we show that $S_a \subseteq A$. Let r be any element in S_a such that $r < a$. Since $a = \sup A$, there exists an element s in A such that $r < s < a$. Then $g(r) <' g(s)$, and $g(s) \leq' a'$. Thus, $r \in A$. If $a \notin \mathbb{Q}_{\mathbb{F}}$, then $S_a \subseteq A$, and therefore, $S_a = A$ in this case. So we assume that $a \in \mathbb{Q}_{\mathbb{F}}$. Then we prove that $g(a) = a'$. If $a' <' g(a)$, then there exists an element b' in $\mathbb{Q}_{\mathbb{F}'}$ such that $a' <' b' <' g(a)$. Since g is a bijection, there exists an element $b \in \mathbb{Q}_{\mathbb{F}}$ such that $g(b) = b'$. Now $a' <' g(b) <' g(a)$ implies that b is an upperbound of A and $b < a$, a contradiction. Thus, $g(a) \leq' a'$. Again if $g(a) <' a'$, then there exists an element $c' \in \mathbb{Q}_{\mathbb{F}'}$ such that $g(a) <' c' <' a'$. Also, there is a $c \in \mathbb{Q}_{\mathbb{F}}$ such that $g(c) = c'$. Therefore, $g(a) <' g(c) <' a'$ which implies that $c \in A$, and that $a < c$. This contradicts the fact that $a = \sup A$. Thus, by the law of trichotomy, $g(a) = a'$. This means that $a \in A$, and therefore, $S_a = A$. Since $S_a = A$, $g(S_a) = g(A) = S'_{a'}$. Therefore, $h(a) = a'$ so that h is surjective.

Step-5. We show that h preserves addition, multiplication and order, that is,

$$h(a + b) = h(a) \oplus h(b),$$

$$h(a \cdot b) = h(a) \odot h(b),$$

and $a < b$ if and only if $h(a) <' h(b)$.

We have already shown that if $a < b$, then $h(a) <' h(b)$. Let $h(a) <' h(b)$. Then $a = b$ is clearly not possible since in this case $h(a) = h(b)$. If $b < a$, then as proved above, $h(b) <' h(a)$, a contradiction. Thus, by the law of trichotomy, $a < b$. Thus,

$$a < b \quad \text{if and only if } h(a) <' h(b).$$

Now since h is order preserving, a subset B of $\mathbb{F}$ is bounded from above in $\mathbb{F}$ if and only if $h(B)$ is bounded from above in $\mathbb{F}'$. Let B be a bounded from above subset of $\mathbb{F}$. Then $h(B)$ is bounded from above in $\mathbb{F}'$. Let $\sup B = a$. We show that $\sup h(B) = h(a)$. Since $x \leq a$ for all $x \in B$, $h(x) \leq' h(a)$ for all $x \in B$, that is, $\sup h(B) \leq' h(a)$. If possible, let $\sup h(B) <' h(a)$, then since h is bijective, there exists an element a' in $\mathbb{F}$ such that $\sup h(B) = h(a') <' h(a)$. Then $a' < a$. Since $h(x) \leq' h(a')$ for all x in B, $x \leq a'$ for all x in B, that is, $\sup B \leq a' < a$, a contradiction. Thus, $h(a) = \sup h(B)$.

Now let a and b be two elements of $\mathbb{F}$. Then

$$\begin{aligned} h(a) \oplus h(b) &= \sup g(S_a) \oplus \sup g(S_b) \\ &= \sup\{g(x) \oplus g(y) : x \in S_a \text{ and } y \in S_b\} \\ &= \sup\{g(x + y) : x \in S_a \text{ and } y \in S_b\} \\ &= \sup\{h(x + y) : x \in S_a \text{ and } y \in S_b\} \end{aligned}$$

since $g(x) = h(x)$ for all x in $\mathbb{Q}_{\mathbb{F}}$; Also, the set $\{x+y : x \in S_a \text{ and } y \in S_b\}$ is bounded from above since $a+b$ is its upper bound. Therefore, $\{h(x+y) : x \in S_a \text{ and } y \in S_b\}$ is bounded from above so that its supremum exists; and its supremum is $h(s)$, where

$$s = \sup\{x+y : x \in S_a \text{ and } y \in S_b\} = \sup S_a + \sup S_b = a+b.$$

Thus, $h(a) \oplus h(b) = h(s) = h(a+b)$.

Finally, we show that $h(a) \odot h(b) = h(a \cdot b)$ for all $a, b \in F$. Let a and b be two elements in $\mathbb{F}$. We have the following cases.

Case (i). $a > 0$, and $b > 0$. Let $P_a = \{x \in \mathbb{Q}_{\mathbb{F}} : 0 < x \leq a\}$, and $P_b = \{x \in \mathbb{Q}_{\mathbb{F}} : 0 < x \leq b\}$. Then

$$\begin{aligned} h(a) \odot h(b) &= \sup g(S_a) \odot \sup g(S_b) \\ &= \sup g(P_a) \odot \sup g(P_b) \\ &= \sup\{g(x) \odot g(y) : x \in P_a \text{ and } y \in P_b\} \\ &= \sup\{h(x \cdot y) : x \in P_a \text{ and } y \in P_b\} \end{aligned}$$

since, $g(x) = h(x)$ for all x in $\mathbb{Q}_{\mathbb{F}}$. The set $\{x \cdot y : x \in p_a \text{ and } y \in P_b\}$ is bounded from above; $a \cdot b$ being its upper bound. Therefore, the image set $\{h(x \cdot y) : x \in P_a \text{ and } y \in P_b\}$ is also bounded from above; and its supremum is $h(s)$, where

$$s = \sup\{x \cdot y : x \in P_a \text{ and } y \in P_b\} = \sup P_a \cdot \sup P_b = a \cdot b.$$

Thus, $h(a) \odot h(b) = h(s) = h(a \cdot b)$.

Case (ii). If $a = 0$, or $b = 0$, then $a \cdot b = 0$. Then $h(a) = \bar{0}$ or $h(b) = \bar{0}$, where $\bar{0}$ is the zero element of $\mathbb{F}'$. Hence, $h(a) \odot h(b) = \bar{0} = h(a \cdot b)$.

Case (iii). $a < 0$ and $b > 0$. Then $-a > 0$, and by case (i),

$$h(-a \cdot b) = h(-a) \odot h(b).$$

Now we use the fact that $h(-a) = -h(a)$ (Here we have used the fact that the infimum of a bounded set exists in complete ordered field). Then we get

$$-h(a \cdot b) = h(-(a \cdot b)) = h(-a \cdot b) = h(-a) \odot h(b) = (-h(a)) \odot h(b) = -h(a) \odot h(b).$$

Therefore, $h(a \cdot b) = h(a) \odot h(b)$.

Remark. In 1936, Tarski introduced real numbers axiomatically. The interested reader is referred to Tarski [45].

EXERCISE

1. Prove the following statements.

   (a) Let $\epsilon > 0$ be a real number. Then there exists a rational number x such that $0 < x < \epsilon$ in $\mathbb{R}_C$.

(b) Let $\epsilon > 0$ be a real number. Then there exist infinitely many rational numbers x such that $0 < x < \epsilon$ in $\mathbb{R}_C$.

(c) For any real number $\xi > 0$, there exists a positive rational number z such that $z < \xi$.

(d) For any real number $\xi > 0$, there exist infinite number of rational numbers z such that $z < \xi$. Also there are infinitely many rational numbers z such that $\xi < z$.

(e) If $\xi > 0$ is a real number, then there exists an $n \in \mathbb{Z}_+$ such that $\frac{1}{n} < \xi$.

(f) Between any two real numbers, there exist infinitely many rational numbers.

(g) Let ξ be an irrational number, and x be a rational number. Then $\xi + x$ and ξx are both irratonal numbers.

(h) Between any two real numbers, there are infinitely many irrational numbers.

2. Show that in a complete ordered field $\mathbb{F}$ every non-empty subset which is bounded from below has its infimum in $\mathbb{F}$.

3. Show that a monotonically decreasing sequence $\langle a_n \rangle$ of real numbers which is bounded from below converges, and it converges to the greatest lower bound of the set $\{a_1, a_2, \ldots\}$ of points of the sequence $\langle a_n \rangle$.

Appendix A

Definition of Sets as Cumulative Type Structures

Most working mathematicians agree that the collections whose members are numbers (of some fixed sort) are sets: and further the sets whose members are such sets, and the sets of those etc. This kind of argument thinks sets as cumulative type structures in which we begin by recognising some simple type of collections as sets and then construct from these in stages more and more complex type of sets. The following collections are sets.

Level 0: Some individuals (about which we presuppose nothing).

Level 1: All collections whose members are individuals.

Level 2: All collections whose members are in level 0 or 1.

At each level we take all collections whose members are in earlier levels. Thus, if the individuals are natural numbers, the collection $\{1, 2\}$ occurs at level 1, the collection $\{1, \{2\}\}$ occurs are level 2. $\{1, \{2\}, \{1, \{2\}\}\}$ at level 3, and so on.

Note that every set occurs at every level after the one at which it is introduced, and it is taken the same at every occurrence. Two obvious questions arise with this definition. What shall we take as individuals? And, how far do the levels go? A cumulative type theory of sets has been developed by taking no individuals at all at level 0. (This is, in fact equivalent to taking the empty set as the only set at level 1). There is a justification for this: if set theory is to be the foundation of mathematics, it is unsatisfactory to include some individuals about which we know nothing: and the sets which we get from the empty set will turn up any way, and so they have to be studied whether there are other sets or not. The levels in the above definition continue indefinitely, since the sets in a given level are used to form sets in the next level. For details on cumulative type set theory, the reader is referred to Drake [15].

Appendix B

Bibliography

A. A. Franklin, Y. Bar-Hillel and A. Levy, *Foundations of Set Theory*, North Holland, Amsterdam (1973).

A. Brown and C. Pearcy, *An Introduction to Analysis (GTM)*, Springer Verlag, New York (1995).

A. G. Hamilton, *Numbers, Sets and Axioms*, Cambridge University Press, UK (1983).

A. Tarski, *Introduction to Logic and to the Methodology of Deductive Sciences*, Oxford University Press, New York (1965).

B. Ram, *Metric Spaces*, Vinayak Publications, Delhi (2005).

B. Rotman and G. T. Kneebone, *The Theory of Sets and Transfinite Numbers*, Oldbourne, London (1972).

C. G. C. Pitts, *Introduction to Metric spaces*, Oliver and Boyd, Edinburgh (1972).

C. W. Burrill, *Foundation of Real Numbers*, McGraw-Hill, New York (1967).

C. W. Dodge, *Numbers and Mathematics*, Prindle, Weber and Schmidt Inc., Boston (1975).

E. Kreyszig, *Introductory Functional Analysis with Application*, John Willey Sons, New York (1978).

E. Mendelson, *Introduction to Mathematical Logic*, D. Van Nostrand Company Inc., Princeton (1964).

E. T. Copson, *Metric Spaces*, Cambridge University Press, UK (1968).

G. Bachman and L. Larski, *Functional Analysis*, Academic Press, New York (1966).

G. Birkoff, *Lattice Theory*, A.M.S. Colloqium Publ. XXV, New York (1981).

G. Cohen, *A Course in Modern Analysis and its Applications*, Cambridge University Press, UK (2003).

G. D. Barru, *Measure Theory and Integration*, New Age International (P) Limited, New Delhi (1981).

G. F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill (1963).

G. Gratzer, *General Lattice Theory*, Birkhauser Verlag, Basel (2003).

G. L. Cain, *Introduction to General Topology*, Addison Wesley Publishing Company, Massachusetts (1994).

H. B. Enderton, *Elements of Set Theory*, Academic Press, New York (1977).

I. N. Herstein, *Topics in Algebra*, Vikas Publishing House (Xerox-Corporation), Delhi (1975).

J. Dieudonne, *Foundations of Modern Analysis*, Academic Press, New York (1964).

J. G. Hocking and G. S. Young, *Topology*, Addison-Wesley, London (1961).

J. I. Nagata, *Modern General Topology*, North Holland Publishing Company, Amsterdam (1968).

J. R. Munkers, *Topology*, Prentice Hall of India (1991).

K. Saxe, *Beginning Functional Analysis*, Springer (UTM), New York (2002).

L. V. Ahlfors, *Complex Analysis*, McGraw-Hill, New York (1966).

L. W. Cohen and G. Ehrlich, *The Structure of Real Number System*, D. Van Nostrand Company Inc., New York (1963).

M. A. Armstrong, *Basic Topology*, Springer Verlag, New York (1983).

M. A. Khamsi and W. A. Kirk, *An Introduction to Metric Spaces and Fixed Point Theory*, John Wiley, New York (2001).

M. Mursaleen, *Elements of Metric Spaces*, Annmaya Publishers, New Delhi (2005).

M. A. Qazi, The mean value Theorem and Analytic Functions of a Complex Variable, *J Math. Anal. Appl.* (324) 30–38 (2006).

N. Bourbaki, *Theory of Sets*, Hermann, Paris (1968).

P. K. Jain and K. Ahmad, *Metric Spaces*, Narosa, New Delhi (1993).

P. R. Halmos, *Finite Dimensional Vector Spaces*, D. Van Nostrand Company, New York (1958).

P. R. Halmos, *Naive Set Theory*, Van Nostrand Reinhold Company, New York (1960).

R. B. Reisel, *Elementary Theory of Metric Spaces*, Springer Verlag, Berlin (1962).

R. Drake, *Set Theory*, North Holland Publishing Company, Amesterdam (1974).

R. Engelking, *General Topology*, Warszawa (1977).

R. R. Goldberg, *Methods of Real Analysis*, Blassdell Publishing Company and Oxford and IBH Publishing Company, New Delhi (1978).

S. Kumaresan, *Topology of Metric Spaces*, Narosa, New Delhi (2005).

W. J. Pervin, *Foundations of General Topology*, Academic Press, New York (1964).

W. P. Eames, *The Elementary Theory of Numbers, Polynomials and Rational Functions*, Oldbourne, London (1967).

W. Rudin, *Principles of Mathematical Analysis*, McGraw Hill International Book Company, Auckland (1985).

Y. F. Liu and S. Y. T. Lin, *Set Theory*, An Intuitive Approach, Houghton Mifflin Company, Boston (1974).

Index

Absolutely convergent, 82–83
Absolute value, 29, 36, 42, 336
Absolutely closed metric spaces, 236
Addition, 12, 22, 26–32, 35–38, 41–43, 62, 68–70, 74–75, 77–79, 240, 242–243, 295, 307, 310, 313, 334, 338, 340, 342–344
Adherent point, 125–126, 129, 158, 166, 187
Algebra, 2–3, 12, 38, 44, 307–311, 313
Algebraic real numbers, 53–54, 259
Archimedean, 33, 36, 39, 340–341, 343
Ascoli-Arzela theorem, 267, 270
Axiom of choice, 54–59, 135

Baire's category theorem, 228–232
Banach fixed point theorem, 324
Banach space, 209, 241–243
Bernstein polynomial, 305–307
Bolzano lemma, 287
Boundary, 127–129, 181, 187, 285
Bounded metric, 92, 107, 115, 144–145, 150–151, 157, 176, 197, 214, 249, 255–256
Bounded sets, 105–109, 141, 149–152, 236–237, 258, 273

Cantor set, 145–148, 259, 292
Cantor conjecture, 64
Cardinal numbers, 60–64
Cauchy filter, 244
Cauchy sequence, 33–34, 39, 82, 193, 208–216, 218, 220, 222–226, 229, 235–239, 244–245, 256, 266, 277–278, 314, 320, 333–336, 339, 341–342
Chordal metric, 93–94
Closed mappings, 190–192
Closed set, 118–120, 122–124, 126, 130, 136–137, 141, 145–146, 148, 162, 168–169, 172, 178–180, 183–184, 186, 190–192, 216, 218–221, 230, 234, 236, 242, 248, 255, 257–258, 261, 264–265, 273–275, 278, 280–283, 285, 290, 292, 296, 298, 312, 315–316, 324
Closure of a set, 121–125, 162, 221
Cluster point of a sequence, 156–158, 252
Commutative algebra, 308–310
Compact metric spaces, 246–249, 300
Compactness of products, 258, 260
Complete metric spaces, 208–216, 246, 255–256
Complete subspaces, 218–219
Completeness of $\mathbf{R}_C$, 339–342
Completeness of products, 218–219
Completion of a metric space, 222–227
Complex plane, 44, 87, 93–94, 127, 171, 182, 193, 210, 242, 261, 279, 297, 312
Components, 289–291, 299–300, 302
Connected sets, 282–285, 289
Connectedness of products, 289
Continuity at a point, 19–20
Continuous everywhere and differentiable nowhere, 276, 316–317
Continuous mappings, 167–172, 178–180, 183–184, 187, 193, 195–207, 238, 240–243, 246, 260–263, 265, 273, 276–277, 286–287
Continuum hypothesis, 61, 63–64
Contraction, 320–322, 324–326, 328, 330
Convergence in products, 159–162
Convergence of a real sequence, 82, 153
Convergence of sequence, 82, 86, 153, 155, 160, 193–194, 333, 339
Convex set, 80, 295, 299
Countably compact, 264–265

Dense set, 129–133, 135, 139, 148, 176–177, 179, 187, 228–229, 266, 275, 316
Denumerable, 49–53, 275, 291, 315
Derived set, 121, 123
Diameter of a set, 124
Disconnected sets, 282
Discrete metric, 87, 103, 107, 122–123, 125–127, 129–130, 132, 135, 149, 153,